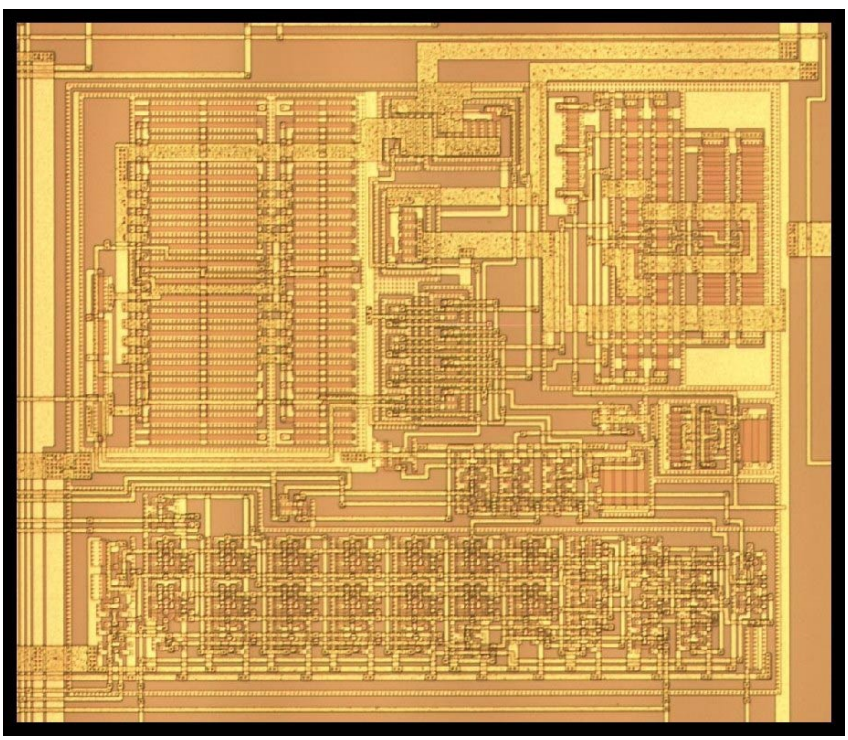
CMOS Circuit Design for Biomedical Telemetry

Gudnason, Gunnar; Bruun, Erik

Publication date:
2001

Document Version
Publisher's PDF, also known as Version of record

CMOS CIRCUIT DESIGN FOR BIOMEDICAL TELEMETRY



Gunnar Gudnason

This thesis was submitted towards the Ph.D. degree at the Technical University of Denmark, Ørsted•DTU, on August 6, 2001.

Document version history

Version 1: LaTeX run August 4, 2001.
Initial submission run.
Times Roman, {`oneside,12pt,a4paper`} book class option.

Version 2: LaTeX run October 25, 2001.
First official print version with fewer typos and reduced page size.
Times Roman, {`twoside,12pt,a4paper`} book class option.
Magnification factor `\mag=800`.

Version 3: LaTeX run November 29, 2001.
Second printing with typographical corrections and minor editing
in response to the opponents' comments.

ABSTRACT

Implanted electronics have a long history, reaching back over 40 years to the first totally implanted battery-powered pacemaker. During this time, electronic implants have become ever more commonplace, but the use of integrated circuit technology for such devices has only become prevalent during recent years.

This thesis presents the results of work done on the design of analog integrated circuits for use in biomedical telemetry systems, specifically for Functional Electrical Stimulation (FES) and for nerve signal sensor applications. The systems are powered by an externally generated electromagnetic field, and data communication to and from the implanted devices is made possible by modulating this field. While this type of telemetry link eliminates the need for wires to the implant or an implanted battery, it sets strict limitations on the power use of the device.

We have designed integrated circuit solutions for many of the problems associated with implanted devices, and present here circuits which improve in several ways over previously published designs, in functionality and level of integration. As an additional constraint, our designs are all implemented in plain CMOS technology, while most other designs rely at least partly on elements available in more specialized (and expensive) technologies. We have succeeded in implementing building blocks for a complete implanted transceiver IC which relies only on an absolute minimum of off-chip components.

The system blocks include circuits for the complete power supply path, including conversion of the RF carrier to a DC supply voltage, overvoltage protection and internal voltage regulation. In addition to being fully integrated in CMOS, these supply path blocks improve on previous designs in performance, functionality and/or power consumption. In addition to the power supply management, we have designed new receiver types which are adapted to the special requirements of mixed power/data transmission. Two complete implantable chips are presented, which highlight the design tradeoffs and optimizations applied to the design of CMOS implantable system chips.

ÁGRIP

Ígræddar rafrásir eiga sér tiltölulega langa sögu, þar sem meira en 40 ár eru liðin frá því að fyrsti rafhlöðuknúni hjartagangráðurinn var tekinn í notkun. Á þeim tíma sem liðinn er frá þeim atburði hafa ígrædd rafeindataæki orðið sífellt algengari lækningatæki, en notkun samrása til þessara tækja hefur ekki orðið almenn fyrr en á seinasta áratug.

Ritgerð þessi greinir frá niðurstöðum rannsókna sem framkvæmdar hafa verið varðandi hönnun hliðrænna samrása fyrir læknisfræðileg fjarkönnunartæki, nánar tiltekið fyrir það sem á ensku nefnist *Functional Electrical Stimulation* og fyrir fjarskynjun taugamerkja. Þessi ígræddu tæki fá alla raforku sína og stýrimerki með rafsegulbylgjum frá sendi sem liggur utan líkamans. Með þessari gerð sambands við ytri stjórnstöð er ekki þörf fyrir leiðslur í gegnum húðina, en kröfur til orkusparnaðar aukast jafnframt vegna lélegrar nýtni.

Hér verður greint frá lausnum sem fundist hafa á mörgum þeim vandamálum sem setja ígræddum raftækjum skorður, og sagt frá rásum sem eru að mörgu leyti eru betri en þær

sem áður hafa birst, meðal annars með tilliti til fjölhæfni og orkusparnaðar. Þar að auki hefur verið notuð einföld CMOS framleiðslutækni, andstætt fyrr birtum lausnum sem nota flóknari og dýrari framleiðslutækni. Tekist hefur að hanna rásarhluta sem hægt er að setja saman í fullkomið ígræðanlegt viðtæki, og sem jafnframt notar lágmarksfjölda íhluta utan kísilflögunnar.

Lýst er rásum fyrir öll þrep í meðferð raforkuflutnings frá ytri sendi, meðal annars rásum til að umbreyta orku á formi rafsegulbylgna við háa tíðni í nothæft form, rásum fyrir yfispennuvörn og fyrir spennureglun. Þessar rásir standa áður birtum lausnum að ýmsu leyti framar, meðal annars með tilliti til afkastagetu, fjölhæfni og orkunotkunar. Ennfremur er sagt frá hönnun viðtækja sem eru sérstaklega aðlöguð kröfum sem stafa frá því hvernig burðarbylgjan er notuð til að flytja bæði afl og upplýsingar. Að lokum eru teknar fyrir tvær kerfislausnir sem leggja áherslu á það hvernig megi hámarka afköst og nýtni ígræddra kerfisrása innan þeirra takmarka sem þeim sett eru.

RÉSUMÉ

Implanteret elektronik har efterhånden en lang historie, der strækker sig over 40 år, fra den gang den første implanterbare batteridrevne pacemaker blev taget i brug. Elektroniske implantater har siden da vundet frem, og er i dag blevet ganske almindelige.

Denne afhandling præsenterer resultaterne af et udviklingsprojekt omfattende design af analoge kredsløb for biomedicinske telemetrisystemer, især systemer til Funktionel Elektrisk Stimulation (FES) og måling af nervesignaler. Disse systemer får deres energiforsyning fra et eksternt elektromagnetisk felt, og dataoverførsel foregår ved modulation af dette felt. Fordelen ved denne type telemetri er at elektriske ledninger ind og ud af kroppen undgås men den effekt der er til rådighed for implantatet bliver stærkt begrænset.

Vi har udviklet integrerede kredsløb der løser mange af problemstillingerne inden for implanteret elektronik, og der vises kredsløb der på adskillige måder udmærker sig i forhold til andre publicerede løsninger. Alle de heri viste kredsløb er implementeret i simpel CMOS teknologi, hvorimod de fleste andre løsninger baseres på brugen af avanceret procesteknologi. Vi har implementeret byggeblokke for en komplet implanterbar transceiver som støtter sig til et minimalt antal eksterne komponenter.

De forskellige systemblokke dækker over kredsløb der udfører alle opgaver inden for behandling af forsyningsspændingen, inklusiv konvertering af RF signalet til DC, overspændingsbeskyttelse og intern spændingsregulering. Foruden at være implementeret i CMOS har disse forsyningsblokke en bedre ydeevne og funktionalitet og et lavere effektforbrug end ældre løsninger. Ud over dem er der designet nye receivertyper der er specielt egnede til de forhold der opstår ved blandet effekt og data transmission. To komplette chipdesign gennemgås for at vise de forskellige optimeringer og kompromisser der indgår i design af implanterbare CMOS systemchips.

CONTENTS

1
INTRODUCTION

Electronic stimulation of human tissue for medical purposes has its roots in the first cardiac pacemakers. The exact date for the first use of a portable implanted pacemaker depends on the definitions of "portable" and "implanted", but some worthy candidates were taken into use around 1960. The art of neurostimulation, where electrical signals are applied to nerves in order to evoke a desired response, evolved in the early 1970s from pacemaker technology. For a long time, these were built from discrete electronics components, and later on, using general purpose integrated circuits. During the last decade or so, the focus has increasingly been on applications of custom integrated circuits for biomedical devices and all the new functionality they can provide. Microelectromechanical Systems (MEMS) have also found a plethora of applications in biomedical devices, but the focus has been aimed more at processing technology development than on circuit design.

The research results presented here are the result of a cooperation between the Department of Information Technology (now Ørsted•DTU) at the Technical University of Denmark and the Center for Sensory-Motor Interaction (SMI) at Aalborg University. Much research has been carried out at SMI regarding the functioning of the nervous system, and the special discipline within the boundaries of this science which is relevant to this work is functional electrical stimulation (FES) or functional neural stimulation (FNS). The objective of FES is to generate artificial nerve signals, and by the correct application of these signals, restore functionality which has been lost due to accident or disease. Several neural stimulators have been built at SMI for stimulation of the motor system, and our work has benefited from the expertise gained in their development and use.

Space restrictions in an implanted device increase the effective cost of every additional component, in terms of usability and reliability. Every off-chip connection which must be added, because of a system function which cannot be integrated on the chip itself, contributes to the reduced reliability of the system. While circuits which reside completely on an IC die can be protected quite well against the harsh environment in the body, the connections to external components can be fragile. One of the main goals of this project has therefore been to find a way to integrate completely on silicon many of the common functions found in implants.

The requirements towards electronics in implants are quite severe in many ways, and many of them have traditionally been met by resorting to exotic IC technologies or external

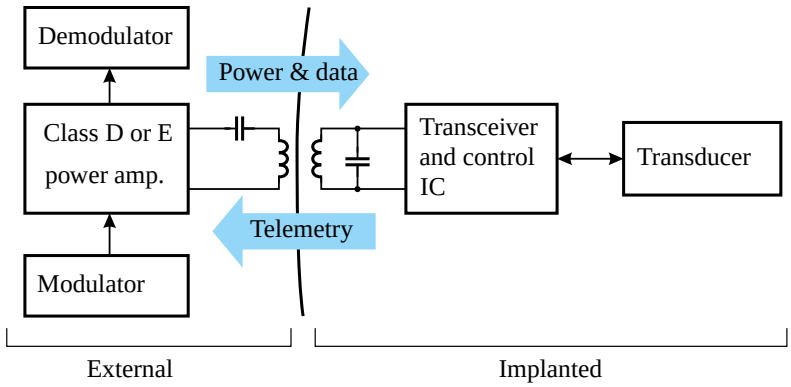


Figure 1.1: *A block diagram showing the main part of an implanted stimulation/measurement system. The power and data is transmitted into the implant and telemetry data is extracted either by load modulating the external field or by active transmission.*

components. There are however many things to be said for plain CMOS technology. The rapid development of CMOS processing technology, while driven mainly by the needs of the digital industry, helps the integration of complex biomedical systems on a single chip. The main disadvantage of the technology evolution is the ever decreasing supply voltage, but as long as the maximum voltage is at least a few threshold voltages, it is still possible to create reasonable mixed analog/digital systems. The second goal of this research has been to implement as much of as possible of the implant circuits in CMOS.

To avoid fixed electrical connections through the skin into the body, or as is the rule in pacemakers, implanted batteries which must be replaced periodically, the systems presented here are designed for use with wireless power and data transmission. The wireless power transmission is implemented in practice through the use of a so-called inductive link, which is basically a weakly coupled coreless transformer. A model of a typical arrangement is shown in figure 1.1. The primary winding of the transformer is driven by a power amplifier which induces a voltage on the implanted secondary. Typical operating frequencies are in the range 1–10 MHz, but examples exist of systems which use higher or lower frequencies. The power supply for the implant is obtained by rectification and filtering of the received radio-frequency signal. The low power efficiency of this link is one of the main driving forces for low-power design in inductively coupled implants. Digital data transmission to the implant is made possible by modulating the RF carrier signal, and reverse transmission can be implemented by varying the load seen by the secondary of the transformer. This type of load shift keying (LSK) is however only possible if the coupling in the transformer is not too small.

The diagram shows two system components in the implanted part, the transceiver and control IC, and a transducer. This transducer can in principle be anything, and is not necessarily on a separate chip. In the case of electrical stimulation it would typically be an electrode connecting an on-chip current source to the nerve. In a nerve signal sensor system it could be a separate electrode and amplifier combination.

The inductive link is the subject of chapter 2. It is a key component of the system and

its characteristics influence most of the interface design of an implantable chip. The power conversion circuit, whose task it is to extract the system's supply power from the RF carrier, must be compatible with the receiver antenna design. The design of receiver circuits for a shared data/power link is also influenced to a large degree by the link.

The link circuit is usually represented as a collection of lumped R, L and C elements, which is sufficiently accurate to describe the electrical circuit-level behavior. The connection between the external transmitter and the implanted system can then be described completely by a single coupling coefficient k. Such a circuit description does however have a severe disadvantage from the link designer's point of view, namely that there is no connection between the circuit description and the physical link realization. We have therefore developed in chapter 2 several expressions which provide a connection between the physical dimensions and the electrical behavior. These expressions are by necessity based on a few approximations, because of the complex dependence of mutual and self-inductances on the geometry. Experimental results did however prove that the expressions were quite accurate.

Several transmitters were designed in the course of this project, and chapter 2 also includes a description of transmitter architectures. A successful class D transmitter is described, and the influence of the transmitter design on efficiency and modulation methods is described.

Chapter 3 treats the design of receivers for links which share data and power transmission. The design requirements for this type of link are unique in at least one respect, which is the input signal range. The input signal amplitude is similar to the supply voltage, and can actually exceed the supply range in the positive and negative directions. The option of using a separate receiver coil does exist but it should be avoided if possible because of the added interconnect complexity in the implant. The shared nature of the link means that the selected modulation method must take the power transfer into account. The best power transfer efficiency is obtained by sending a constant unmodulated carrier, but in order to transfer data, this carrier must obviously be modulated. A common method used to secure the power for the implant while still leaving room for data transfer is to send the data in bursts, with an unmodulated carrier in between. This places additional demands on the receiver design.

An overview of previous receiver designs is given in chapter 3, followed by new receiver designs developed by us. The design criteria were that the receivers should be compatible with burst mode data transfer, low modulation index amplitude shift keying (ASK) and adaptive detection levels. In addition, the receivers are designed for a minimum of supporting logic, especially without the DSP methods which characterize modern digital receiver designs. The omission of a DSP was justified partly because of the savings in supply current, but chiefly because the required design effort was too large for the available resources, and the research value of a traditional DSP solution is minimal.

The on-chip power supply quality of implanted ICs is discussed in chapter 4, and several novel supply management circuits are presented. The supply voltage generation of inductively coupled systems is one area where external components or advanced processing steps have traditionally been used. This regards specifically the rectification of the RF carrier and overvoltage protection, which have previously been carried out on-chip by the use of floating p-n junctions and zener diodes. Since these circuit elements are not available in plain CMOS, we have developed active circuit solutions which can perform the same tasks, and are fully integrated in a CMOS technology. We present a CMOS bridge rectifier which does not suffer from parasitic effects and an active overvoltage protection circuit with a very large

shunt current range. In addition, we examine on-chip voltage regulators and their application in reducing power consumption and providing isolation between analog and digital parts of an integrated circuit.

A self-contained system chip must also generate various support functions internally, such as reference voltages, bias signals and a digital power-on reset (POR). A bandgap voltage reference which can be implemented in a CMOS process is presented in chapter 5, and we show a new power-on reset circuit. The POR circuit uses very little static power, and can be implemented without any advanced analog process features, such as high-ohmic polysilicon. Design tradeoffs are also investigated for the degenerate current mirror structure, or constant-g_m cell, which is often used in CMOS bandgap references and bias circuits.

Chapter 6 shows two examples of system design for implantable devices, and emphasizes the system-level requirements which are specific to inductively coupled implants. These requirements relate mainly to the nature of the power supply, the characteristics of the data transmission and limitations in interconnect capabilities. The first example is an implantable neural stimulator chip, whose purpose is to generate stimulation pulses in response to commands transmitted by an external controller. The second example is a general transducer chip which can be used in conjunction with the generation or measurement of nerve signals. The main purpose of the transducer chip is to exploit the possibilities offered by the spatial separation of components in a neural implant.

2
LINK DESIGN

The design of the inductive link between an external transmitter and control unit and an implanted electronic device must be based on considerations of performance factors. One of these is the power consumption, which in a mobile system must be low enough that the battery change intervals are acceptable to the user. We can regard the power dissipation or loss as taking place in three locations, the transmitter power stage, the link air interface and inside the implant. No actual power dissipation takes place in the path between the transmitter and the receiver but since the unfocused power transfer and the weak coupling of the parts lead to a very low link efficiency, it must be included. The task of the transmitter is to set up at the location of the implant a modulated magnetic field which is strong enough to power the implant. Because of the inefficiency inherent in this arrangement the load placed by the implant upon the transmitter is much smaller than the other loss mechanisms in the system, the transmitter operated almost independently of the receiver. It is therefore useful to consider the transmitter as an isolated system and measure its performance by a yardstick which considers only efficiency in producing the remote field. The only real losses which need to be considered in the transmitter design are then the losses in the transmitter itself and the resistive losses in the primary LC circuit.

A large amount of work has been carried out on the design of coils for inductive links [37, 27], and the results of that work will not be repeated here. Much of that work has however focused on closely coupled links where the implanted device is relatively large and shallowly implanted. The coupling coefficients are therefore larger than we expect to obtain here and the loading of the primary by the secondary is more visible. A more realistic design goal is to maximize the transferred power subject to limitations in the size and quality of the implanted coil. A wide range of materials has been used in the construction of implanted coils, from hand-wound copper wire inductors to on-chip planar spiral coils made with standard metal layers [17, 44] or more exotic metallization layers [9]. We shall give an example of a planar substrate coil implemented with Au thick-film tracks on an alumina substrate.

Performance specifications of the link like power transfer, range and bandwidth are best calculated using a circuit-level description of the link. There are however a few circuit elements which can only be derived form a physical description of the system, namely the inductances of the coils and the mutual inductance or the coupling coefficient. Inductances can be calculated using analytical approximations or semi-empirical models, and we shall give

examples of those for some simple coil geometries. Another possible method is to simply build the coils and measure their inductance but this is only practical for wire coils. The fabrication of inductors on a substrate is usually too expensive and time-consuming for casual experimentation. The best option in that case is to use three-dimensional electromagnetic field simulators to obtain the inductances and loss factors for a given geometry. Most such tools can simulate inductances quite accurately, but their ability to calculate inductor losses is lacking, especially with respect to high-frequency losses in silicon substrates [74, 8]. Since the operating frequency of inductive links is only on the order of 10 MHz and the dimensions are larger than in RF on-chip inductors, the losses are mostly due to bulk resistance and current crowding in the conductors. Various numerical tools can therefore be used to calculate the inductances and resistances with sufficient accuracy.

The range of the link is closely connected to the power transfer since the range is the distance at which the transferred power falls below the requirement of the implant. As will be shown below, the power transfer is a strong function of the distance, so the range is nearly constant regardless of the power consumption of the implant. An expression for the range can be derived based on this and some other reasonable approximations. There are two main factors which contribute to the difficulty in the determination of the power transfer. One is the complex dependence of the mutual inductance on geometrical details and the other is the nonlinearity in the power converter. The mutual inductance can be calculated by making a few simplifying assumptions about the geometry, or by using a three-dimensional field solver. The power converter is the subject of section 4.2. An alternative method is to use a field simulator to obtain the mutual inductance and then use a circuit simulator to calculate the power transfer. This gives highly accurate results which are valid for a single set of operating parameters, but it does not provide a tool for circuit synthesis.

2.1 Circuit-level link description

The behavior of the inductive link between the external and internal circuitry can be treated at two different abstraction levels, the electromagnetic field description and at the circuit level. These two abstractions are in a way complementary, yet it can be difficult to make the transformation from one to the other. The electromagnetic description is based on the geometry and physical composition of the link elements, but does not give direct information about the way in which electrical signals propagate through the transmission channel implemented by the link. The circuit-level description can on the other hand be used to derive channel characteristics, but key parameters of the circuit elements like inductances, parasitics and coupling coefficients can only be extracted from the electromagnetic field description. This parameter extraction is a nontrivial task, and some possible procedures will be outlined in the next section. In this section we will focus on the effect of the circuit parameters on a few important performance points, namely the inbound power transfer, the channel bandwidth and the outbound data transfer capability. Outbound data transfer is in the last case implemented through passive load modulation, and an estimate of the feasibility of this type of data transfer can be based on the sensitivity of the external (primary) circuit on the operating point or configuration of the internal (secondary) circuit. Active outbound data transmission schemes which bypass the power link have been described in [68, 29], but these are not relevant to the characteristics of the link.

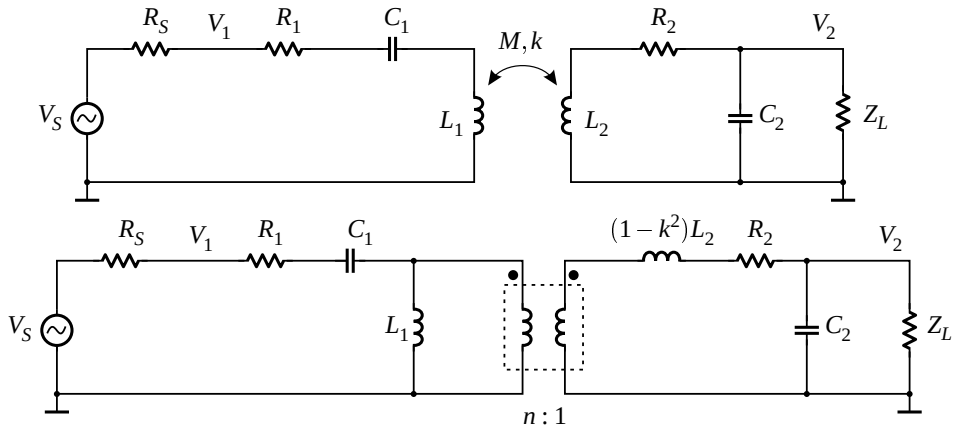


Figure 2.1: *A linear model of the weakly coupled inductive link. The nonlinear power converter attached to the secondary is omitted. Part (b) shows an equivalent circuit where the weakly coupled transformer in (a) has been replaced with an equivalent ideal transformer with two uncoupled inductors.*

2.1.1 Transformer model

The link can be treated along the same lines as a lossy, weakly coupled transformer [2, 99]. The contributions to the losses in the "transformer" stem from the series resistances in the inductors, while core losses are of course absent. The link inductors are tuned to the operating frequency for better efficiency, so the circuits corresponds to a double-tuned transformer. The primary circuit is in most cases tuned in series to provide a low-impedance load to the driving transmitter at the resonant frequency. The secondary is on the other hand almost invariably a parallel LC circuit to better drive a nonlinear rectifier load. The high output impedance of the parallel circuit is the best match for a rectifying power conversion circuit, which for all practical purposes looks like a rectifier connected to a stiff voltage source. The nature of the power converter load is treated in more detail in section 4.2. Figure 2.1 shows a working model of the transformer/link with the parasitic resistances of the inductors shown as R_1 and R_2. When the secondary circuit is analyzed as an isolated circuit it can be useful to transform the series resistance into an equivalent parallel resistor with a simple series-to-parallel transformation $R_{2p} = Q_2^2 R_{2s}$. Such a transformation is however only valid at a single frequency and to maintain generality as long as possible we refrain from transforming the resistance. Figure 2.1(b) does however show a transformation of the two coupled inductors into an equivalent magnetic circuit [2]. This transformation is exact in the absence of core losses, which is obviously the case here. The equivalent turns ratio for the ideal transformer is

$$n = \frac{1}{k}\sqrt{\frac{L_1}{L_2}} \tag{2.1}$$

The relation between the coupling coefficient k, the inductances L_1 and L_2, and the mutual inductance M should be made clear. The mutual inductance is a physical parameter which depends on the geometry of the link, while the coupling coefficient is a circuit pa-

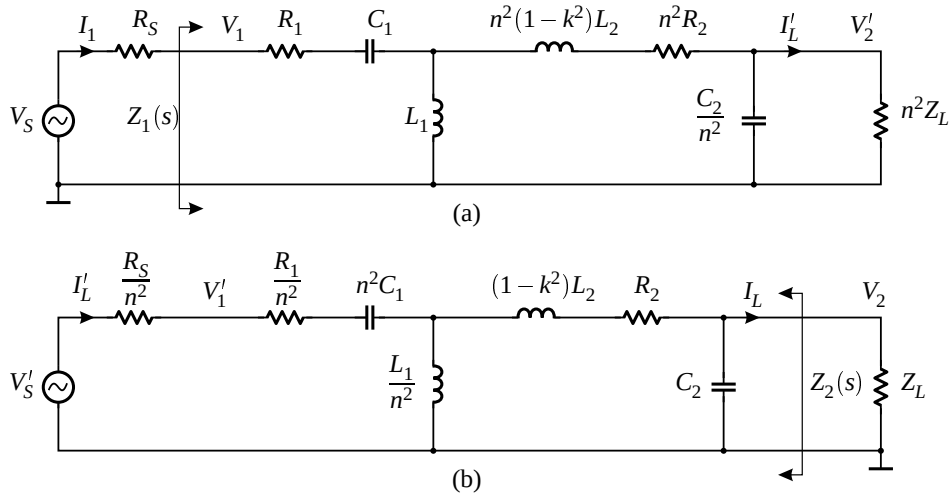


Figure 2.2: *The link circuit referred to the primary side in (a) and to the secondary side in (b).*

rameter used to describe the effect of the mutual inductance in the circuit abstraction. The relation between the two is given by

$$M = k\sqrt{L_1 L_2} \tag{2.2}$$

In the next section we show some examples where three-dimensional field simulators have been used to calculate the self- and mutual inductances of specific link geometries. The physical inductances obtained in that way provide the link between a geometrical description and the circuit abstraction of the system.

The characteristics of the link can be observed from two points of view. At the primary side, the input impedance defines the drive requirements for the transmitter. An analytic expression for the input impedance also shows the effect of the secondary circuit on the primary, and can be used to gauge the observability of load variations in the secondary circuit. This can tell us whether load modulation is a practical proposition for outbound data transfer for a given link coupling. The properties of the link seen from the secondary side are useful to calculate the power transfer and the transmission channel characteristics for inbound data.

2.1.2 Transfer functions

By using the equivalent transformer circuit in figure 2.1(b) one can refer everything to the primary or the secondary side. With the turns ratio defined as in (2.1) and in figure 2.1, an impedance Z across the secondary output is equivalent to an impedance n^2Z across the primary input. This simple rule leads to the circuits shown in figure 2.2(a) and (b) where everything has been referred to the primary and secondary side respectively. These can be used to derive expressions for the most interesting link performance parameters, namely the power or current transferred to the load, the output requirement for the signal source, the

bandwidth of the link and the sensitivity of the primary input impedance on the secondary load.

The voltages and currents obtained by putting the circuits in figure 2.2 through the transfer function grinder need to be transformed back to their real values afterwards. Transformed values are identified here with an apostrophe, so we have $V_2' = nV_2$, $I_2' = I_2/n$, $V_s' = V_s/n$ and so on.

These double-tuned links have fourth-order bandpass transfer functions, with two pairs of complex conjugate poles. The primary and secondary link sides are in practice always tuned to close but not quite equal frequencies to reduce the sensitivity of the link gain to coupling variations [37]. This fact can be used to justify some assumptions which can simplify the analysis. However, oversimplification can hide features of the circuit dynamics which may be important, so it is better to simplify as the last stage in the derivation. The load seen by the signal generator can be shown to be

$$Z_1(s) = \frac{A(s)}{B(s)} \tag{2.3}$$

where

$$\begin{aligned} A(s) = {} & n^2(R_2 + Z_L) + s\Big[\frac{L_1}{k^2} + n^2R_2C_2Z_L + n^2C_1R_1(R_2 + Z_L)\Big] \\ & + s^2\Big[n^2L_1C_1(R_2 + Z_L) + C_2Z_L\frac{L_1}{k^2} + C_1R_1\big(\frac{L_1}{k^2} + n^2R_2C_2Z_L\big)\Big] \\ & + s^3\Big[C_1C_2R_1Z_L\frac{L_1}{k^2} + n^2L_1C_1\big((1-k^2)L_2 + R_2C_2Z_L\big)\Big] \\ & + s^4L_1C_1C_2Z_Ln^2(1-k^2)L_2 \end{aligned} \tag{2.4}$$

and

$$B(s) = sn^2C_1(R_2 + Z_L) + s^2C_1\big(\frac{L_1}{k^2} + n^2R_2C_2Z_L\big) + s^3\frac{L_1}{k^2}C_1C_2Z_L \tag{2.5}$$

The dependence of this impedance function on the secondary load Z_L is needed to determine the observability of load variations at the transmitter. Clearly, this dependence cannot be deduced by inspection of (2.3), and differentiation of the impedance with respect to Z_L is unlikely to bring enlightenment. The expression can however be used for numerical analysis of the link behavior. An approximate treatment of the load modulation effect on the primary is given in [102], where the primary and secondary are tuned to a common frequency and the secondary parasitic resistance R_2 is absorbed in the load Z_L. In addition, the load is assumed to have a constant reactive component which can be absorbed into the capacitance C_2, leaving a real load R_L. In that case, the voltage amplitude across the primary inductor depends on the load resistance R_L according to

$$V_{L1}(R_L) = \frac{V_S}{\omega C_1(R_S + R_1 + \omega^2k^2L_1C_2R_L)} \tag{2.6}$$

If the load resistance is modulated between the two values R_{L0} and R_{L1}, the modulation index of the inductor voltage is

$$m = \left|\frac{V_{L1}(R_{L0}) - V_{L1}(R_{L1})}{V_{L1}(R_{L0}) + V_{L1}(R_{L1})}\right| = \omega^2k^2L_1C_2\left|\frac{R_{L0} - R_{L1}}{R_{L0} + R_{L1}}\right| \tag{2.7}$$

This is a useful expression because it tells us that the maximum voltage modulation index perceived across the primary inductor is equal to $m_{max} = \omega^2 k^2 L_1 C_2$. The inductances L_1 and L_2 are usually similar, because while the secondary inductor may have a smaller diameter it is often practical to use more turns in the secondary. The assumption $L_1 = L_2$ leads to the result that the maximum modulation index is equal to the coupling coefficient squared k^2. Small coupling coefficients therefore require a considerable effort in the design of the sensor circuit.

The effect of load modulation can be detected in several places in the primary circuit, but using an amplitude sensor across the primary inductor is a common choice [31, 102]. There are also examples where a current sensor has been used to measure the current in the primary circuit through a transformer, where the primary circuit in the link is also a single primary turn in this transformer [107].

The source impedance seen by the secondary load influences the design of the power conversion circuit. The exact expression of the source impedance derived from the diagram in figure 2.2(b) is

$$Z_2(s) = \frac{C(s)}{D(s)} \tag{2.8}$$

where

$$\begin{aligned} C(s) = {} & n^2 R_2 + s\big(L_1 + n^2(1-k^2)L_2 + n^2 R_2 C_1 (R_S + R_1)\big) \\ & + s^2\Big[n^2 C_1\big(R_2 L_1 + (1-k^2)L_2(R_S+R_1)\big) + L_1 C_1 (R_S + R_1)\Big] + s^3 n^2 L_1 C_1 (1-k^2) L_2 \end{aligned} \tag{2.9}$$

and

$$\begin{aligned} D(s) = {} & n^2 + s n^2\big(C_1(R_S+R_1) + R_2 C_2\big) \\ & + s^2\Big[n^2\big(L_1 C_1 + (1-k^2)L_2 C_2 + C_1 C_2 R_2 (R_S + R_1)\big) + L_1 C_2\Big] \\ & + s^3\Big[n^2\big(R_2 C_2 L_1 C_1 + (1-k^2)L_2 C_2 C_1 (R_S+R_1)\big) + L_1 C_1 C_2 (R_S + R_1)\Big] \\ & + s^4 n^2 (1-k^2) L_1 C_1 L_2 C_2 \end{aligned} \tag{2.10}$$

To simplify the source impedance expression we may however omit the effect of the primary completely, and only take L_2, C_2 and R_2 into account. This is justified because in contrast to the load modulation case, the figure of interest (which is the source impedance in this case) is mainly governed by the near side of the link, and the far side effects are secondary. The output impedance is then that of a simple *RLC* circuit and is equal to

$$Z_2(s) = \frac{R_2 + sL_2}{1 + sR_2 C_2 + s^2 L_2 C_2} \tag{2.11}$$

Close to the resonance frequency the output impedance may be estimated through a series-to-parallel transformation to $Z_2 \approx L_2/(R_2 C_2)$. The validity of (2.11) can be verified by taking the uncoupled limit of (2.8), by letting $k \to 0$ and $n \to \infty$.

The voltage transfer function from the driving source to the secondary output can be calculated in a similar way, but the result is a rather complex function which is not directly useful for further analysis. If the driving frequency is a common resonance frequency for

both sides of the link it can be shown [27] that the voltage transfer function of the link at that frequency is

$$A(\omega_0) = \frac{V_2}{V_S} = \left[\left(\frac{C_1(R_1 + R_S)}{C_2(R_{2p}||R_L)k} + k \right) \sqrt{\frac{C_2}{C_1}} \right]^{-1} \tag{2.12}$$

This expression is based on the assumption that the secondary load Z_L is resistive and that the driving frequency is slightly offset from the resonant frequency resulting in a reflected load which is real for all k. This last trick has little importance for weakly coupled links since the detuning of the primary is small compared to its bandwidth. By inspection of (2.12) we see that the function has a maximum for a certain coupling coefficient k_{crit}. This critical coupling point can be exploited to obtain a link transfer function which is insensitive to small displacements of the coils relative to each other. This mode of operation is however only practical for links with strong coupling, or in approximate terms, when the diameter of the implanted coil is equal to or larger than the depth of implantation. For weak coupling the $1/k$ term in (2.12) dominates, and we may approximate the transfer link gain as

$$A_w(\omega_0) = k \frac{(R_{2p}||R_L)}{R_1 + R_S} \sqrt{\frac{C_2}{C_1}} \tag{2.13}$$

Not surprisingly, the secondary voltage is proportional to the coupling coefficient, or to the mutual inductance. It should be noted that this is assuming a linear secondary load, which is definitely not the case when a rectifier-based power converter is used. An equivalent resistive load which dissipates the same power is sometimes used to model the circuit [27, 102] but this is rather misleading. The assumption is that the power conversion circuit is a rectifier with a linear resistive load on the other side of the rectifier, while in reality it looks more like a rectifier with a stiff voltage source on the other side (see section 4.2). The power conversion circuit therefore acts as a voltage clamp which restricts the secondary voltage to a range which depends on the converter topology.

2.1.3 Range

The range of the link can be difficult to determine because of the nonlinear secondary load. The equivalent linear load can be used to calculate some parameters of the link, but the approximation is only valid at a single combination of transmitter power and link distance. A useful measure of the link range can be obtained by calculating the range at which the secondary voltage amplitude is equal to the desired implant supply voltage (plus a fitting number of rectifier voltage drops, to be more precise). The practical range is somewhat shorter because some power must be transferred into the load, so some forward bias is necessary across the nonlinearity. The error in this distance estimate is however relatively small because of the strong dependence of the field on the distance. An advantage of treating the range in this way is that it can be obtained by analyzing a linear circuit, and without relying on linear approximations of nonlinear elements. The only load on the secondary circuit is then the parasitic R_2 associated with the inductor, and it can be transformed by a series-parallel transformation without loss of generality, since we assume that the transmitter frequency is known.

The procedure is to find the k at which the link gain reaches the input threshold of the power converter, and subsequently translating this coupling coefficient to a physical distance

using the methods in the next section. The converter threshold is a function of two variables, the internal supply voltage of the implant and the voltage drop in the converter. The voltage drop $|V_D|$ depends on the converter topology and is roughly two rectifier forward drops for a bridge converter and one rectifier drop for most other types. As we shall see in section 4.2 these rectifiers are not necessarily based on p-n junctions so we cannot in advance determine $|V_D|$.

If we assume that the range is determined by the k for which the secondary voltage reaches the converter threshold $V_2 = |V_D| + V_{DD}$, we have

$$A_w V_S = V_S \frac{kR_{2p}}{R_1 + R_S} \sqrt{\frac{C_2}{C_1}} = |V_D| + V_{DD} \tag{2.14}$$

or

$$k = \frac{(|V_D| + V_{DD})(R_1 + R_S)}{V_S R_{2p}} \sqrt{\frac{C_1}{C_2}} \tag{2.15}$$

This is the minimum coupling coefficient needed to provide the full supply voltage for the implant. This k can be translated to a physical separation d using expressions developed in section 2.2.3.

2.2 Electromagnetic view

The complex geometries involved in the analysis of coupled coils in inductive links make it impossible to derive general expressions for the link characteristics, like power transfer and induced secondary voltage. If we assume that the lumped-element model of each side of the link is known, the only parameter which must be found to completely determine the electrical behavior is the coupling coefficient k, or equivalently the mutual inductance M. By restricting the possible geometries to a few special cases, it is possible to derive analytic expressions for these quantities.

2.2.1 Field calculations

As a starting point for calculations of magnetic fields created by transmitter coils, we can use the law of Biot and Savart, which in integral form reads as

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0 I}{4\pi} \oint_C \frac{d\mathbf{s} \times \mathbf{R}}{R^3} \tag{2.16}$$

where I is the current flowing in a current loop C, $\mathbf{s}$ is a unit vector along the current path and $\mathbf{R}$ is the vector from the path element to the measurement point. The voltage induced in a secondary circuit can be found using Faraday's law of induction, which states that the voltage U induced in a circuit C due to a time-varying magnetic induction $\mathbf{B}$ is

$$U = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{u} \tag{2.17}$$

where S is any surface linking the circuit and $\mathbf{u}$ is a unit vector normal to the surface. We follow here the convention of calling the field $\mathbf{B}$ the magnetic induction, while the term *magnetic field* is reserved for the vector field $\mathbf{H}$ which includes the effects of matter.

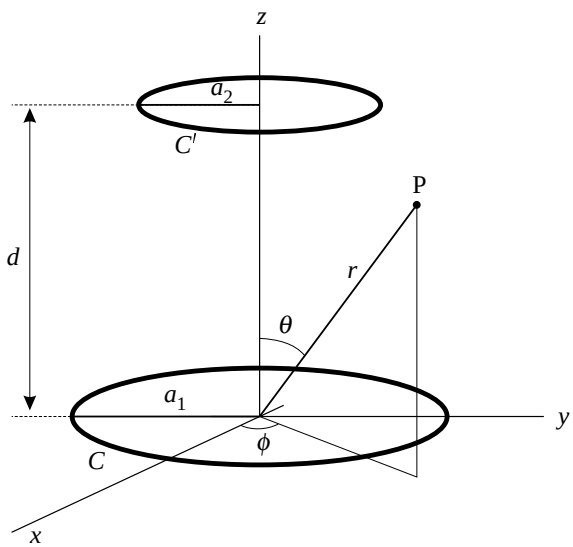


Figure 2.3: *An overview over the link geometry with two coaxial circular current loops and the definition of the coordinates used.*

There exists a special case for which the Biot-Savart law can be applied directly to obtain an exact expression of the magnetic induction. On the axis of a circular current loop with current I and radius a, at a distance r from the plane of the loop, the field is

$$B = \frac{\mu_0 I a^2}{2(a^2 + r^2)^{3/2}} \tag{2.18}$$

directed along the axis. The choice of the symbol r for the separation is to maintain consistency with other results below which are presented in spherical coordinates. The field intensity is strongly dependent on the distance, falling as the third power of the distance far from the loop. For this reason, the range of inductive links is quite limited, at least when they are used for power transfer. The $1/r^3$ dependence can also cause problems when the link distance is reduced since the transferred power may rise above the level which the receiving device can handle, and destroy the device. This happened a few times while we were testing devices described in this work, and the power of the transmitter had to be degraded by inserting large power resistors in series with the transmitter coil. This underlines the fact that inductively coupled devices must have a robust mechanism for dissipating excess power if they are to survive.

The magnetic induction from a circular current loop can be calculated at an arbitrary point using (2.16), but this leads to an expression involving elliptical integrals which must be evaluated numerically. There are some specific regions in the space around the loop where the symmetry leads to a simplification of the expression. Using spherical coordinates (r, θ, ϕ) where ϕ is the azimuthal angle, it can be shown [52] that the magnetic induction close to the axis from a loop with radius a may be expanded as the series

$$B_r = \frac{\mu_0 I a^2 \cos\theta}{2(a^2 + r^2)^{3/2}} \left[1 + \frac{15 a^2 r^2 \sin^2\theta}{8(a^2 + r^2)^2} + \dots \right] \tag{2.19}$$

and

$$B_\theta = -\frac{\mu_0 I a^2 \sin\theta}{4(a^2+r^2)^{5/2}}\left[2a^2 - r^2 + \frac{15a^2r^2\sin^2\theta(4a^2-3r^2)}{8(a^2+r^2)^2} + \dots\right] \tag{2.20}$$

The azimuthal field component B_ϕ is zero because of the axial symmetry. The special regions where the expression can be simplified are close to the axis ($\theta \ll 1$), close to the center of the loop ($r \ll a$) and far from the loop ($r \gg a$). The last case is the most interesting here, since there exists an exact expression for the on-axis field and an inductive link is rarely operated with concentric coils. Far from the loop, the field reduces to that of a magnetic dipole $m = \pi I a^2$,

$$\begin{aligned} B_r &= \left(\frac{\mu_0 I a^2}{2}\right)\frac{\cos\theta}{r^3} \\ B_\theta &= \left(\frac{\mu_0 I a^2}{4}\right)\frac{\sin\theta}{r^3} \end{aligned} \tag{2.21}$$

As a last note on analytic field calculations, a few words of explanation should be added about the meaning of the terms *near field* and *far field* in this context. The expression *far field* in electromagnetic systems is usually reserved for the radiation region where wave propagation determines the field. In that respect, all inductive links operate in effect as near-field antennas, since the combination of antenna size and operating frequency means that the characteristic dimensions are much smaller than the wavelength. For this reason, it is possible to use quasi-static approximations for the derivation of link behavior. The inaccuracy caused by this approximation is no larger than the error caused by other approximations in this treatment. As an example we may take the extreme case of high frequency and large link dimensions. At a frequency of 100 MHz a distance of 10 cm corresponds to 1/30th of a wavelength in free space. Since the ratio of distance to wavelength is far smaller in most cases we can forget about radiation effects with a clear conscience.

2.2.2 Radiated power

Even if radiation effects have little effect on the operation of an inductive link, and the power lost through radiation is small in most cases, compliance with local frequency spectrum regulation may require that the characteristics of the radiated power be known. A radiation loss of 1% for example is insignificant compared to other losses in a transmitter, but if the total power is 10W, the transmitter may be in violation of regulations.

A typical primary coil is circular in shape, so it can be modeled quite accurately as a magnetic dipole. If some other shape is used, higher terms in a multipole expansion can be used to describe the radiator. The dipole term does however dominate at normal link frequencies so the dipole term is sufficient. The magnetic dipole moment of a loop with area A and N turns is $m = INA$. The total radiated power from a magnetic dipole source is [117]

$$P = \frac{\mu_0 \omega^4 m^2}{12\pi c^3} \tag{2.22}$$

where c is the speed of light. The fourth power of the frequency causes the radiated power to increase very rapidly as the frequency is increased. The frequency dependence of the

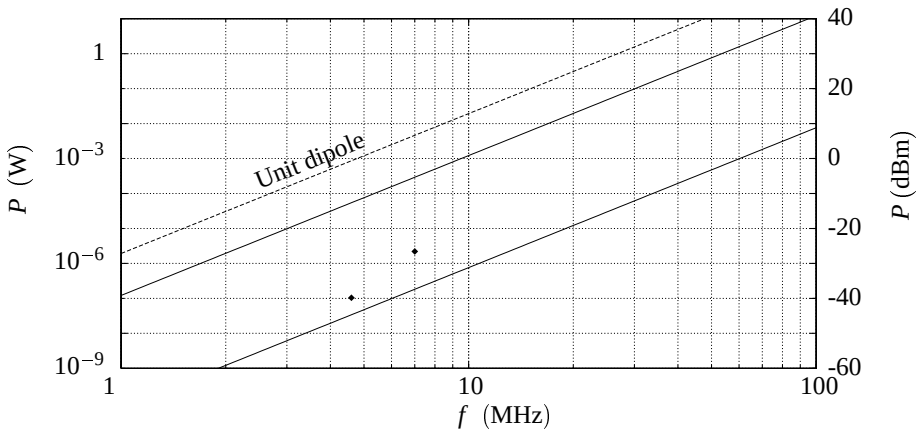


Figure 2.4: *The total radiated power from two different transmitter loops as a function of frequency. The upper trace is for a 10-turn 8 cm diameter coil with a current of 5A, and the lower trace is for a 5-turn 4 cm diameter coil and 1A. Most transmitters fall between these two limits. Some typical operating points for the transmitter described in section 2.4 are also shown. For reference, a unit dipole with I=1 A and an area of 1 m^2 has been added to the graph.*

quadrupole term is ω^6, the sextupole term ω^8 and so on, but their magnitude at normal link frequencies is small enough to be neglected.

A plot of (2.22) is shown in figure 2.4 for two loop antennas which represent the high and low ends of the range of dipole moments used with biomedical implants. The figure is somewhat misleading since the straight lines in the plot do not take into account the difficulty in driving the current loop at high frequencies with a constant AC current. As the frequency increases, the capacitance needed to tune the fixed loop decreases until the parasitic capacitances dominate and the self-resonant frequency is reached. We can conclude that the radiated power at frequencies below 10 MHz is unlikely to cause problems. At higher frequencies or with extremely powerful transmitters, there is however cause for concern, and the radiation question should be evaluated on a case-by-case basis.

2.2.3 Effect of geometry on coupling

The goal of the preceding field calculations was to find a starting point in the evaluation of the effects of geometry on the coupling constant between the primary and secondary antenna. We have an expression for the magnetic induction from a circular current loop at various levels of approximation, and it will now be applied together with Faraday's induction law to find the induced voltage in the secondary circuit.

The special case of coaxial circular current loops deserves to be mentioned because there exists a closed-form expression for the mutual inductance [117, 82]. The mutual inductance

is given by

$$M_{12} = \frac{\mu_0 a_1 a_2}{4\pi} \int_0^{2\pi} \int_0^{2\pi} \frac{\cos(\phi_2 - \phi_1) d\phi_1 d\phi_2}{\sqrt{d^2 + a_1^2 + a_2^2 - 2a_1 a_2 \cos(\phi_2 - \phi_1)}} \tag{2.23}$$

The elliptic integral in the expression does not have an explicit solution so we must resort to other methods to find more useful expressions.

If we assume that the two current loops are approximately coaxial and parallel, meaning that the axial shift is small compared to the separation and the rotation is small, we can use (2.19) for the field at the center of the second loop. If we assume furthermore that the angle spanned by the secondary seen from the primary is small, or equivalently that the diameter of the secondary is small compared to the separation, we obtain the flux Φ by multiplying the field by the area. Small rotations of the secondary can be included by adding a cosine factor to the flux expression. Since one definition of the mutual inductance of two loops is the flux through one divided by the current in the other, we have

$$M_{12} = \frac{\mu_0 \pi N_1 N_2 a_1^2 a_2^2}{2(a_1^2 + d^2)^{3/2}} \tag{2.24}$$

where we have added the possibility of each loop having multiple turns N. The mutual inductance and the coupling coefficient $k = M/\sqrt{L_1 L_2}$ for weakly coupled links therefore follow the same asymptote as the induction field strength, and fall as the cube of the separation.

An expression for the minimum k necessary for the implant to reach its full supply voltage was derived in (2.15). That expression for k can be used together with (2.24) to obtain the maximum range

$$d_{max}^2 = \left[\frac{\mu_0 \pi N_1 N_2 a_1^2 a_2^2}{2\sqrt{L_1 L_2}} \frac{V_S R_{2p}}{(|V_D| + V_{DD})(R_1 + R_S)} \sqrt{\frac{C_2}{C_1}} \right]^{2/3} - a_1^2 \tag{2.25}$$

The meaning of the first term within the brackets can be clarified by pointing out that the inductance of an N-turn circular inductor with radius a is

$$L = \mu_0 N^2 a \left[\ln\left(\frac{8a}{b}\right) - 2 \right] \tag{2.26}$$

where b is the wire bundle diameter [82]. This is only an approximation since the inductance is sensitive to the details of the geometry. If we furthermore assume that the logarithm term is the same for both inductors (a reasonable approximation since the value of the $\ln x$ function is relatively insensitive to changes in the value of the argument), the range becomes

$$d_{max}^2 = a_1 a_2 \left(\frac{C_2}{C_1}\right)^{1/3} \left[\frac{V_S R_{2p}}{(2\ln(8a_1/b_1) - 4)(|V_D| + V_{DD})(R_1 + R_S)} \right]^{2/3} - a_1^2 \tag{2.27}$$

Although formally complex, this equation only depends on simple physical factors. $V_S/(R_1 + R_S)$ is the primary current amplitude and $(V_D + V_{DD})/R_{2p}$ is the secondary current amplitude at the maximum range.

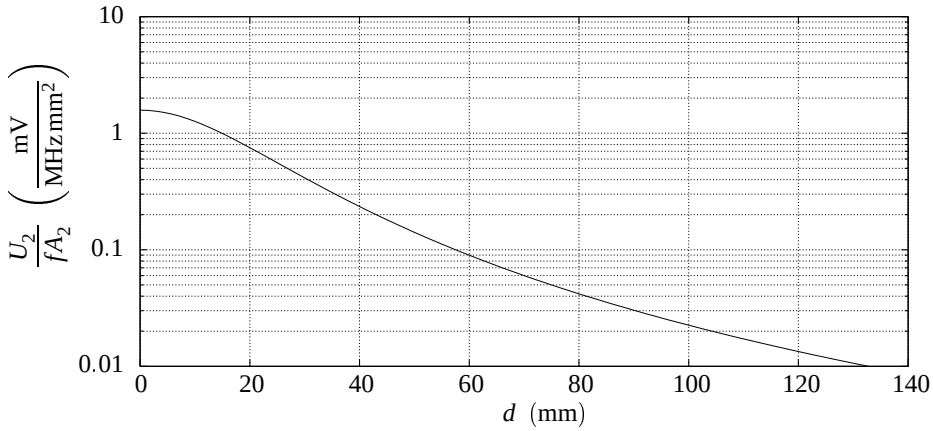


Figure 2.5: *The induced voltage in a circuit loop per MHz for a unit area of 1mm². The curve is calculated for a representative transmitter with a 5-turn 54mm diameter coil and a current amplitude of 1A.*

2.2.4 Induced voltages

The induced voltage in a current loop in the field from the transmitter is interesting for more reasons than just to calculate the voltage in the receiver antenna. All circuit loops in an implanted system, whether they are on a substrate or on-chip, are subject to induced voltages. The size of these loops is presumably much smaller than the spatial variation of the induction field, so the induced voltage can be calculated through Faraday's law by using a fixed value for **B**. If the angle between the field vector and the normal vector of the loop is α, the induced voltage is

$$U_2(t) = -\frac{d}{dt}|\mathbf{B}(\mathbf{r},\mathbf{t})|A_2\cos\alpha \tag{2.28}$$

The field vector **B** can be calculated using the far-field expression for a magnetic dipole. As a more accurate alternative, the field and the induced voltage can be calculated using a 3D field solver. For a simple analytical expression we may however assume that the circuit loop which we are examining is close to the axis of the transmitter coil, and obtain the result

$$U_2(t) = \omega e^{j\omega t}\frac{\mu_0 N_1 I_1 a_1^2}{2(a_1^2+d^2)^{3/2}}A_2\cos\alpha \tag{2.29}$$

The impact on small signals within the system can be considerable if the layout of the circuit is not made carefully. In general, it is best to keep ground and supply loops as small as possible and route signals as close as possible to the rail to which they are referred. Differential signals should be routed with closely spaced tracks and in the case of weak signals and strong fields, some kind of twisting of the track layout may be advisable. The induced voltage is shown in figure 2.5. The fraction of this voltage seen at any point on an implanted integrated circuit depends on the impedance level, but it is clearly large enough to disturb sensitive analog circuits.

As an example we can take a differential signal which is routed across 3 millimeters across a chip with a pitch of 10 µm. With the reference transmitter used in the figure and a

frequency of 7 MHz, operating at a distance of 100 mm, the induced differential voltage has an amplitude of 5 µV. In the case of power supply rails which usually enclose a far greater area, the induced voltage is much larger.

2.3 Coil design

The design of coils for use in an inductive link is based on the desired performance and the restriction which are enforced by either external conditions like size limitations or by the available materials. There are a few general observations which can be made, for example that it is impossible to obtain a large coupling coefficient if the coil separation is not small compared to the coil dimensions. The diameter of the external coil is generally larger than that of the internal coil, because we assume that it can lie flat against the surface of the body. The dimensions and material of the transmitter can therefore for all practical purposes be chosen freely. The main limitations are in the design of the implanted coil, and we shall in the following give a short overview of how these limitations influence the coil design and the operation of the link.

2.3.1 Area versus inductance

The importance of the coil inductance was apparent in the link transfer function calculation of the preceding pages. What was perhaps not as clear was the importance of the coil dimensions for the efficiency of the link. The inductance of a coil with N turns and some linear dimension a scales as $L \propto N^2 a$, while its effective total area scales as $A \propto N a^2$. Equation (2.24) shows that the mutual inductance of the coils is the product of the secondary area and a factor which depends on the primary dimensions and the separation. The secondary inductance does not enter explicitly into the expression. The equation for the range of the link (2.27) also shows that it is primarily the coil diameters and not their inductances which determine the range. The inductances can be said to influence the range through the tuning capacitances in (2.27), but this influence is weak because of the $1/3$ exponent.

2.3.2 Planar inductors

Planar spiral inductors have become a very popular research subject in recent years within RF IC design, but they also have applications in macroscopic circuit design, in PCB inductors and transformers. Because spiral inductors can easily be fabricated on various substrates with good repeatability and accuracy, they are attractive for other uses than on-chip elements. A transmitter coil could for example very well be fabricated by etching the coil pattern in copper on a flexible substrate. Some examples of planar inductors have also been published for microimplants, with the inductor fabricated in a thick electroplated nickel layer on a silicon substrate [9].

In the course of this project, a stimulator chip was mounted on a thick-film substrate with a few discrete components to test a packaging concept. This substrate design is described in greater detail in section 6.1.3, but it is relevant here because the receiver coil was implemented with tracks on the substrate. The approach taken in the design was to use all the

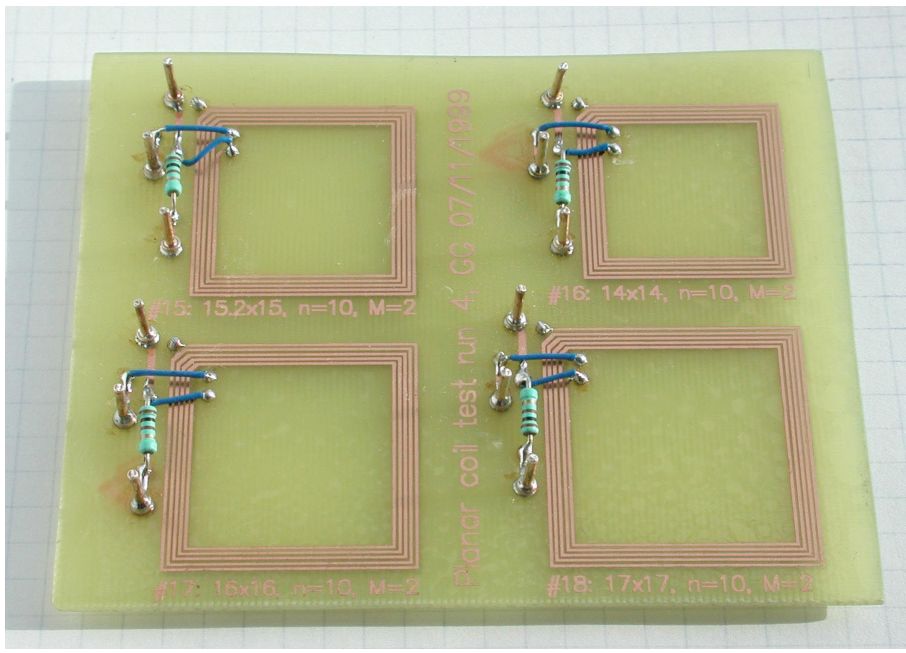


Figure 2.6: *Experimental two-sided planar spiral coils fabricated by etching copper tracks on a FR4 substrate. The dimensions were scaled up by a factor of 2 because the etching lithography did not give the same resolution as a thick-film process.*

available area (about 15×15 mm) for the spiral inductor, and use as many turns as the rest of the components on the substrate allowed.

Exact expressions for the inductance of planar spiral coils do not exist, but there are a few empirical ones which come relatively close to the real value. One example for square spirals is attributed to Wheeler [63],

$$L = \frac{37.5\mu_0 N^2 a^2}{11d - 14a} \tag{2.30}$$

where d is the outer diameter and a is the mean "radius" of the turns. This expression does not take into account the ratio of the turns' width to their pitch, so its accuracy is only moderate. Another method described in [67] which is based on a completely empirical fit to experimental date has proved to give significantly better results. The inductance is given by the monomial

$$L = \beta d_{out}^{\alpha_1} w^{\alpha_2} d_{avg}^{\alpha_3} N^{\alpha_4} s^{\alpha_5} \tag{2.31}$$

where d_{out} is the outer diameter, d_{avg} is the average diameter, w is the track width, s is the track spacing, N is the number of turns and β is a shape-dependent parameter. The α exponents are the fit parameters. It is claimed in [67] that this expression performs as well as a field solver. Although these fit parameters were extracted from data for on-chip coils they have proved correct to within a few percent for PCB-scale spiral inductors.

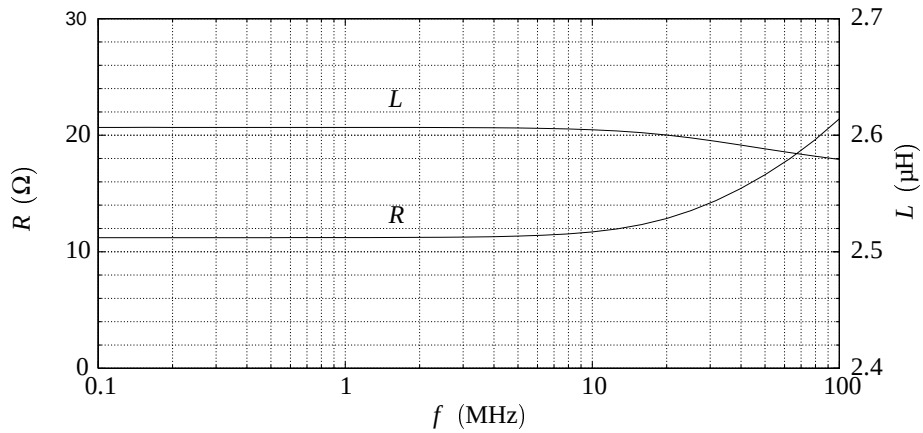


Figure 2.7: *The inductance and resistance of a thick-film substrate coil, calculated using FastHenry.*

The inductance formulas above were however inadequate for our purposes because we wished to use both sides of the thick-film substrate for the inductor. Because the inductance of a multi-layer spiral depends strongly on the layer separation, it is not feasible to distill the inductance into one general empirical expression. To check the inductance before committing the design to fabrication, we made several test runs with identical, but scaled up, inductors on double-sided PCB substrates. An amusing example of this primitive technology is shown in figure 2.6.

Loss mechanisms in integrated inductors are a rather complex affair since they involve not only resistive losses in the conductors but also substrate effects which are difficult to estimate. The case of the thick-film inductor is much simpler since the ceramic substrate is a perfect insulator and the skin depth in the conductor at low MHz frequencies is larger than the conductor cross section. The parasitic resistance is therefore easily found from the bulk resistance of the material.

2.3.3 Field simulations

Because of the limited validity of the explicit inductance expressions for complex geometries we resorted to a three-dimensional field simulator to find the inductances of substrate coils. There are on one hand several commercial software packages which can perform inductance calculation and extract parasitic capacitances and resistances for arbitrary geometries. On the other hand, there are open source packages which are freely available which can perform the same calculation. We have used a program by the name of FastHenry, developed specifically for inductance calculations of complex geometries [56, 57].

Figure 2.7 shows the inductance and parasitic resistance calculated by applying FastHenry to the thick-film substrate coil used for the stimulator chip. The figure shows that the parasitic resistance is constant up to a frequency of about 10 MHz, where skin effects begin to affect the parasitic resistance. The simulated inductance of the coil is 2.61 μH, which is within the measurement uncertainty range for the experimental results (see section 6.1.3).

The field simulator can also be used to find mutual inductances between circuit elements,

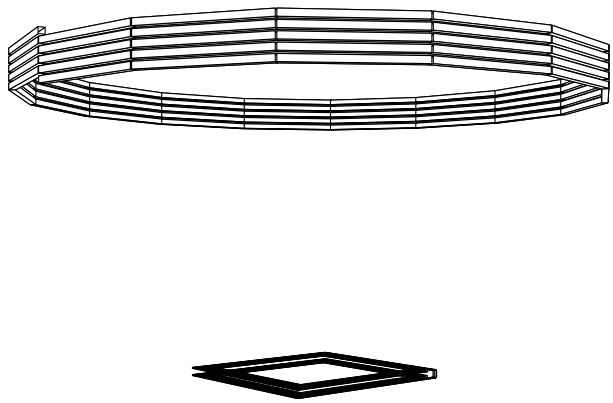


Figure 2.8: *The physical model of the transmitter and receiver coil used for the link coupling coefficient calculation.*

which is very useful to obtain numerical results for the coupling coefficient range achievable with a given set of link elements and coil separation. It is therefore possible to compare different coil geometries quantitatively, eliminating some of the guesswork. We have applied the field simulator to a link consisting of a reasonable transmitter coil and the substrate coil discussed above. The coupling coefficient was calculated for varying receiver coil displacements, resulting in the coupling coefficient map shown in figure 2.9.

2.4 Transmitter design

This discussion of the link has up to now been focused only on the passive elements. In order to be useful, an active transmitter circuit must be incorporated into the link to allow power and data transfer. Several transmitters were designed during this project, not because of their inherent research interest, but to provide the necessary tools to test integrated circuits for implants. Since the transmitter design sets the range of operating parameters available with the link, like power transfer, modulation range and modulation types, an overview of the transmitter is in order.

2.4.1 Architectures

Transmitters used to power inductive links are almost invariably switching types where the active devices are used as approximations of ideals switches. Using the standard classification of power amplifier operating modes, these are class D or E stages.

The difference between class D and E amplifiers lies mainly in the way the switching losses affect the overall efficiency. It is true of both types that the ideal efficiency, disregarding switch losses, is close to 100%. The energy losses during the period while the switches are changing between the on and off states are mainly due to resistive losses in the switch devices, charge lost in the parasitic capacitances and energy stored in the lead inductances. The last term reflects the magnetic energy stored in the device lead inductances which is

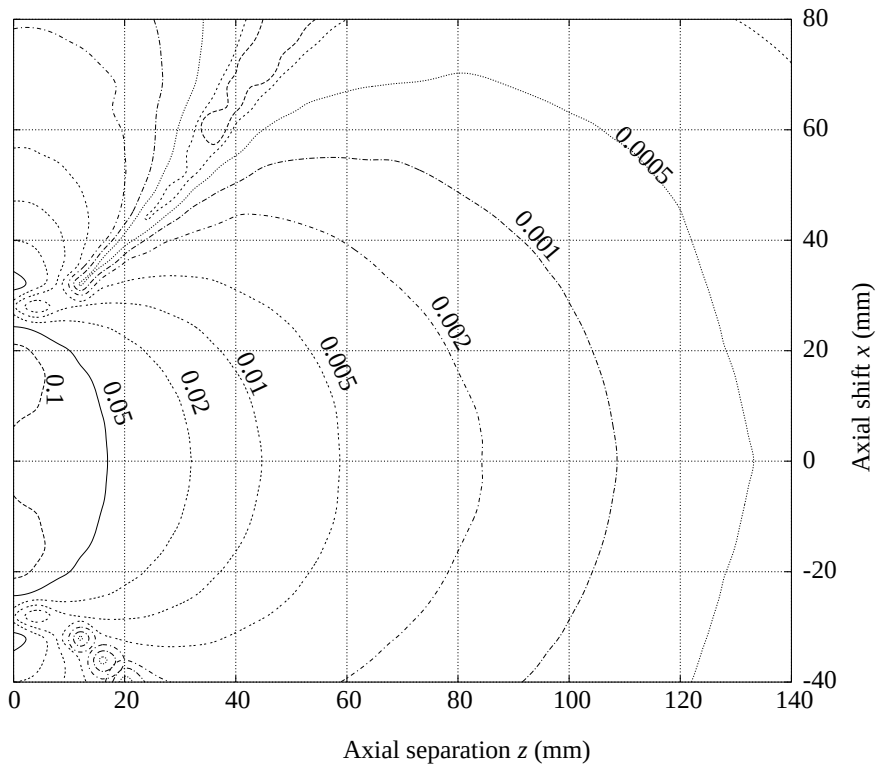


Figure 2.9: *The coupling coefficient k between a 50 mm-diameter circular 5-turn transmitter coil and a rectangular double-sided thick film coil on a ceramic substrate (described in section 6.1.3). The coils are arranged with their axes parallel to the z axis, and the transmitter coil has its center at* $(x, y, z) = (0, 0, 0)$*. The coupling coefficient k is calculated as a function of the separation along z and the relative sideways shift along x. The curves mark contours of constant k. Some of the constant-k contours cross each other because the plotting program had difficulties with abrupt gradient changes.*

dissipated in parasitic resistances when the switch is turned off. This term can be important with standard power device packages at normal link frequencies (ca. 10 MHz), and the problem has in one case led to the design of a special power device package [33]. Class E amplifiers have a practical efficiency advantage over class D types because the active devices switch at zero voltage *and* zero current. It is therefore possible to design a class E amplifier with better efficiency than a similar class D one, since the resistive losses in the switches are smaller. Due to the way in which the class E load is tuned it is also possible to absorb the switch capacitances in desired circuit components.

Class E amplifiers have been used successfully and published by several groups working on implanted electronics [107, 106, 123]. They do however have one disadvantage which makes them less suitable for use in a test tool than their class D counterparts, namely that the load tuning is more intricate. The class D amplifier with a simple LC tuned circuit load

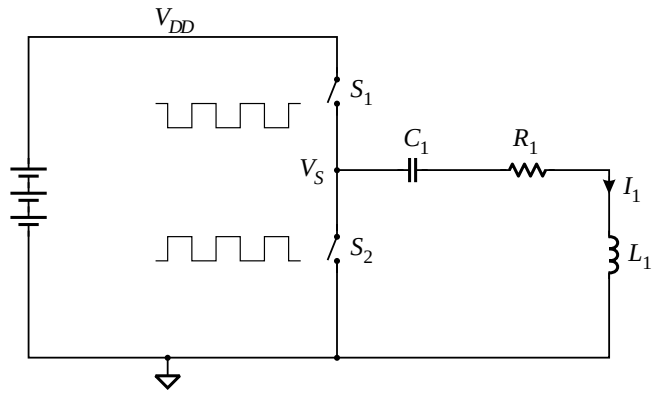


Figure 2.10: *A simplified diagram of a class D amplifier used in the output stage of an inductive link transmitter.*

will function with any frequency and duty cycle as long as the load is tuned to track the frequency changes. It is therefore possible to design a single class D transmitter which can cover a wide range of frequencies (0.5–10 MHz in our case) by changing a single component in the load circuit. An easy way to build the tunable LC load is to use several capacitors in parallel which can be switched in and out of the circuit. The fact that the class D amplifier functions over a wide range of switch duty cycle makes it a simple matter to modulate the transmitter output. If the transmitter is only meant to operate at a single frequency, there is a definite efficiency advantage in the class E design, and that is the preferred choice for portable transmitters operating at a fixed frequency.

The principle of operation of the class D amplifier is shown in figure 2.10. The switches S1 and S2 are driven with complementary signals, and the voltage V_S at their common drive point is a square wave of amplitude V_{DD}, frequency f and duty cycle D. Because of the nonzero transit time of real switches, the switch drive signal overlap should be controlled carefully. The drive point voltage is not a perfect square wave either but we shall approximate it as such to derive a few characteristics of the amplifier.

The tuned load acts as a bandpass filter, suppressing all but the fundamental driving frequency. The amplitude of the fundamental can be found by calculating the Fourier coefficients of the square drive signal. In general, the nth Fourier coefficient ($n > 0$) of a periodic signal with period P is

$$a_n = \frac{2}{P}\int_0^P V_S(t)\cos\left(\frac{2n\pi t}{P}\right)dt \tag{2.32}$$

By choosing the time origin suitably, we find that the general result for a square wave with duty cycle D ($0 \leq D \leq 1$) is

$$a_n = \frac{2V_S}{n\pi}\sin(n\pi D) \tag{2.33}$$

The zeroth term is the only even-order term whose amplitude is nonzero. The fundamental amplitude a_1 is largest for a 50% duty cycle, and the amplitude can be modulated by changing the switch drive duty cycle. The amplitude of the first few odd harmonics amplitude is shown in figure 2.11. The choice of operating point depends on the desired output power

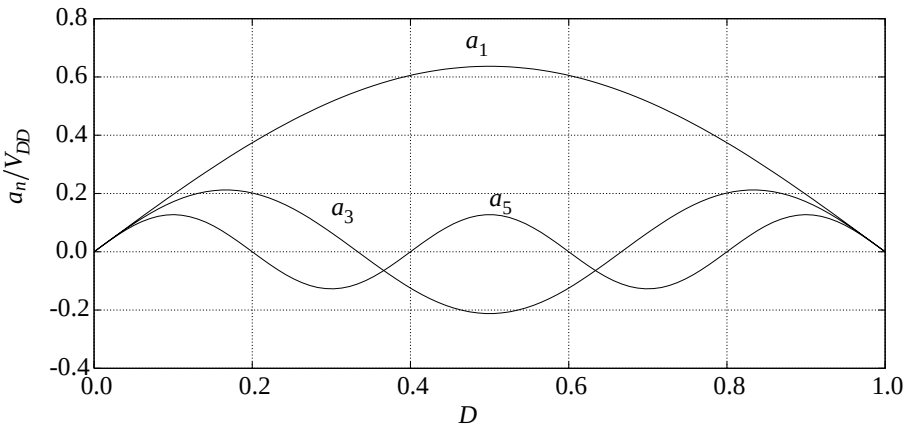


Figure 2.11: *The first nonzero odd-order Fourier coefficients in the expansion of the load driving signal $V_S(t)$ as a function of duty cycle.*

and modulation index. Operation near $D = 0.5$ results in the highest efficiency since the load current is at its maximum and the amplifier losses are nearly independent of the duty cycle. The modulation transfer function is however zero at that point, and duty cycle control cannot be used to modulate the output. A possible alternative modulation method, which can be used if the duty cycle is fixed, is to vary the supply voltage. This method has been used for class E amplifiers [123], but the design of an efficient modulation circuit for the low-impedance supply signal is a challenging task in itself.

If the Q of the load circuit is not very high the amplitude of the harmonics may be important. The figure shows that there is a point $D = 1/3$ where the third harmonic is zero, so the amplitude of the third harmonic into the load may be reduced by choosing this as the operating point.

The current amplitude in the load at the fundamental frequency is $I_1 = a_1(D)/R_1$. The current due to the harmonics can be estimated by using the ratio of the RLC impedance at a frequency $n\omega_0$ to the impedance at resonance

$$\left|\frac{Z(n\omega_0)}{Z(\omega_0)}\right| = \sqrt{1+Q^2\left(\frac{n^2-1}{n}\right)^2} \approx Q\frac{n^2-1}{n} \tag{2.34}$$

The ratio of the harmonics current to the fundamental is then

$$\frac{I_n}{I_1} = \frac{a_n(D)}{a_1(D)}\,\frac{n}{Q(n^2-1)} \tag{2.35}$$

This expression is only an approximation, and is certainly not valid for large n since the increase in the resistive component at higher frequencies due to skin effects was ignored.

The switches in figure 2.10 are implemented with power MOS transistors, which are the preferred device type for this kind of application since they are as close as possible to ideal switches. They do have some nonidealities, namely a large gate capacitance, nonzero switching times and on resistance. The on resistance contributes to the effective resistance

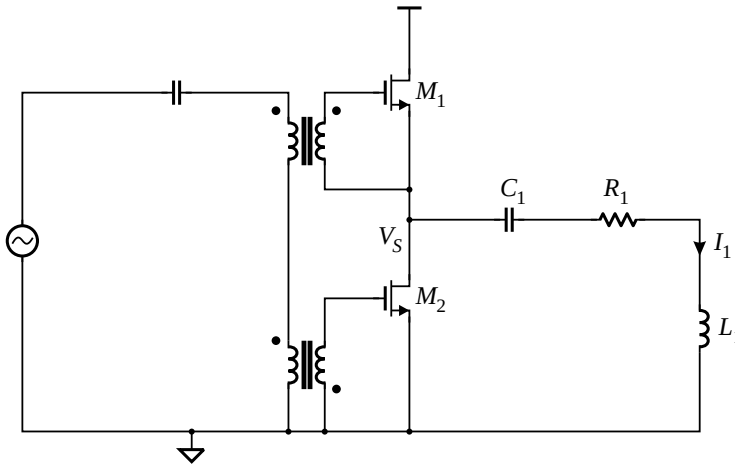


Figure 2.12: *An example of a resonant gate circuit used to reduce the power dissipation in the gate drive. The transformers are wound on toroidal ferrite cores.*

of the tank, so this parameter is important for the efficiency of the amplifier. The gate drive circuit must be capable of handling the rather large gate capacitances of the power devices, which have values in the neighborhood of 1nF. Gate capacitance numbers can be misleading however, both because of the intrinsic nonlinearity of the oxide capacitance and because of the marked Miller effect. Device manufacturers therefore normally quote the total gate charge Q_G needed to turn the device on.

If a wideband driver is used to drive the gates of the devices, the power dissipated per device is approximately $P = fQ_G V_{DD}$. For the IRF520, which is a typical n-channel power MOS, the gate charge is 15nC. The power dissipated in the drive circuit with a supply voltage of 6 V and a frequency of 5 MHz is then 0.45W per transistor. In order to reduce this power dissipation one may use a resonant gate drive for the power transistors, as shown in figure 2.12. This arrangement has the additional advantage that it permits the use of n-channel transistors for both switches. N-channel power transistors generally have a lower 'on' resistance than their p-channel counterparts, just as with integrated circuits. The resonant gate drive does however have a serious disadvantage from our point of view, in that it is difficult to vary the transmitter frequency, and duty cycle modulation cannot be used to control the output power.

2.4.2 Implementations

Figure 2.13 shows the diagram for one implementation of a wideband class D transmitter. This transmitter can operate from a supply voltage of 6–12 V, with a frequency range from 500 kHz up to nearly 10 MHz.

The upper end of the frequency range is limited by the power drivers used for the transistor gates. These are intended for switch-mode power supplies which rarely run at higher frequencies than a few hundred kHz. The power dissipation in the driver is therefore the limiting factor, and a few driver ICs were destroyed during the use of the transmitter.

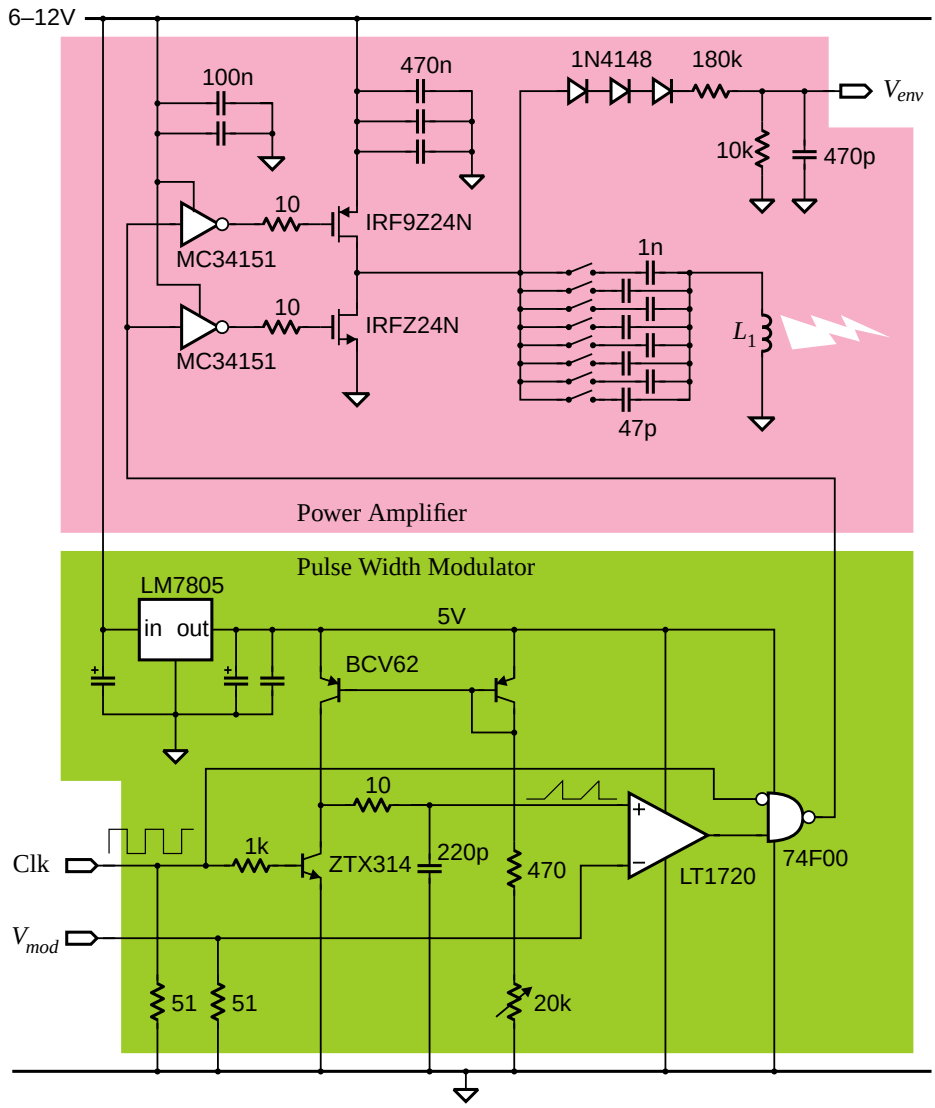


Figure 2.13: *A simplified schematic diagram of the class D transmitter described in the text.*

The center frequency of the load circuit is selected by a switched capacitor bank and interchangeable coils. Because the current amplitude in the load circuit can reach very high values, it can be necessary to share the current between the tuning capacitors by using at least two of them in parallel. The voltage across them can also become very high (up to 600 V in our tests) so the capacitors must be high-voltage types. We used polypropylene capacitors with a rated voltage of 630 V and a low parasitic resistance.

The power devices are IRFZ24N and IRF9Z24N HexFET® transistors from Interna-

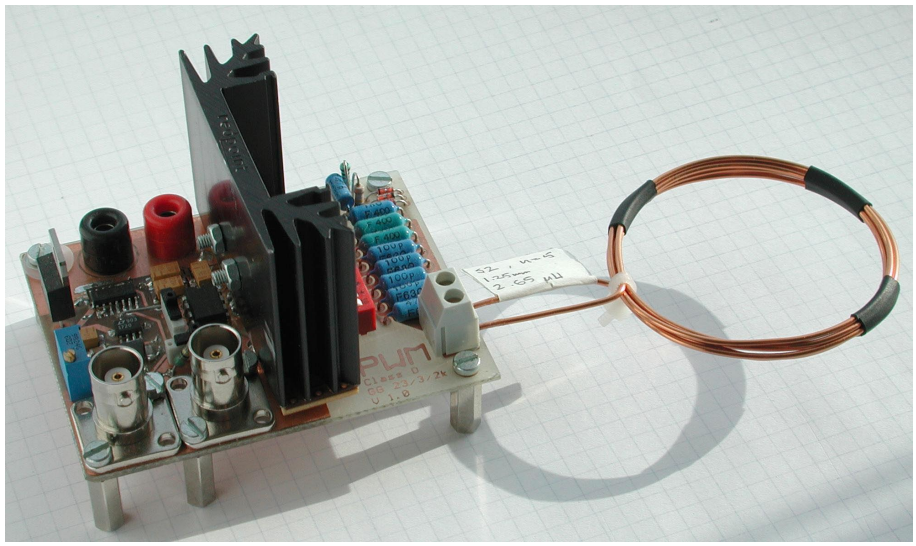


Figure 2.14: *A photograph of the class D transmitter with a 2.6 µH, 55mm diameter coil. The heat sink is much larger than necessary.*

tional Rectifier, with breakdown voltages of 55 V, gate charges of 13nC and on resistances of 0.07 Ω and 0.17 Ω respectively. Power device manufacturers offer large selections of discrete transistors which represent different tradeoff choices in the manufacture of the devices. There is a direct relationship between the on resistance and the gate capacitance as in integrated transistors, and it is best in general to take the lowest breakdown voltage rating which will do the job, as higher breakdown voltages are bought by compromising other performance figures.

The transmitter frequency is set by an external signal source, and the modulation is controlled through a modulation input signal and a PWM circuit. The pulse width modulator is based on the generation of a triangular waveform and comparison thereof with the modulation signal input. These two signals are compared in a very fast LT1720 comparator which has a typical propagation delay of 4.5ns. The nand gate was included to fix the duty cycle at 50% in case the modulation signal falls below the bottom of the triangle signal.

The transmitter was implemented as far as possible with SMD devices to reduce parasitic inductances and ground loops. The fast transients occurring for example in the triangular wave generation circuit can cause ringing and false triggering of the comparator. In addition to making the PCB tracks as short as possible, it was necessary to include small resistors in several places to damp oscillations. A large number of fast decoupling capacitors were also included because of the large current transients. Only a few representative ones are shown in the diagram. Preliminary tests without decoupling capacitors showed supply bounce voltages in excess of 1 V.

The power transistors were mounted on a large heat sink as shown in figure 2.14 since the power dissipation was not known beforehand, and some of the previous transmitter circuits dissipated a lot of heat in the power devices. However, due to the low on resistance of the

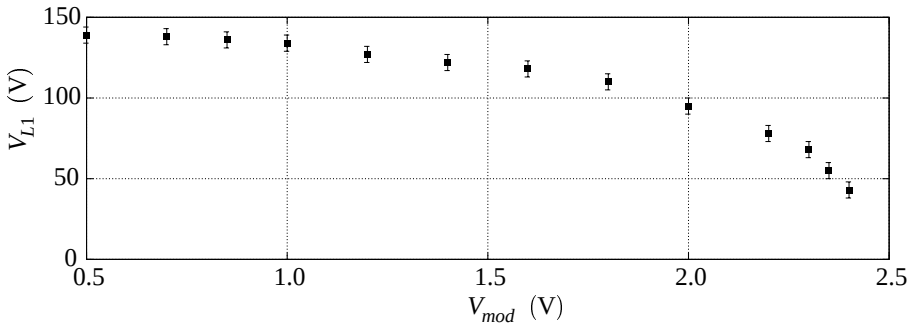


Figure 2.15: *Test results for the modulation of the transmitter output. The frequency was 5.3MHz, the supply voltage 7V, and the coil inductance 4.5µH. The inductor was degenerated with a 4.7Ω resistor to limit the output power. The vertical axis shows the inductor voltage amplitude, and not the RMS value.*

transistors and the efficient topology, the transistors do not become noticeably warm and a much smaller heat sink could have been used.

A simple envelope detector consisting of a few diodes and an RC filter circuit was included in the transmitter. This envelope can be used to measure the output amplitude, either for use in a feedback control of the output as in [31], or for detection of load modulation. Several diodes were used in series because of the low breakdown voltage of fast signal diodes.

2.4.3 Performance

To give an idea of the performance of the amplifier, a few representative results are included here regarding the efficiency, the modulation range and the envelope detector characteristics. Figure 2.15 shows an example of the pulse width modulation effect, and figure 2.16 shows the transfer function for the envelope detector in the output stage of the transmitter. The envelope detector output is a nicely linear function of the inductor voltage amplitude despite the nonlinear elements in the detector.

The last two graphs represent an attempt to estimate the efficiency of the transmitter at two operating frequencies. The load in this context must be the parasitic resistance in the LC circuit, consisting the coil parasitics and the switches' on resistance. These are not easily measured, especially since the coil parasitics are frequency-dependent because of the skin effect and current crowding. The "load resistance" was therefore measured by sweeping the transmitter frequency with a fixed LC load, and using the Q of the LC tank to find the resistance. The equivalent tank resistance and the tank current is then used to calculate the power dissipated in the load.

The graphs show that the efficiency comes quite close to 100%, although the uncertainty in the load resistance determination suggests that the efficiency is not necessarily that good. We did however not have access to equipment which could have given more accurate results. Many of the components used in the transmitter were old, inefficient ICs like the LM78-series regulator and the 74F-series logic, which use more current than more modern

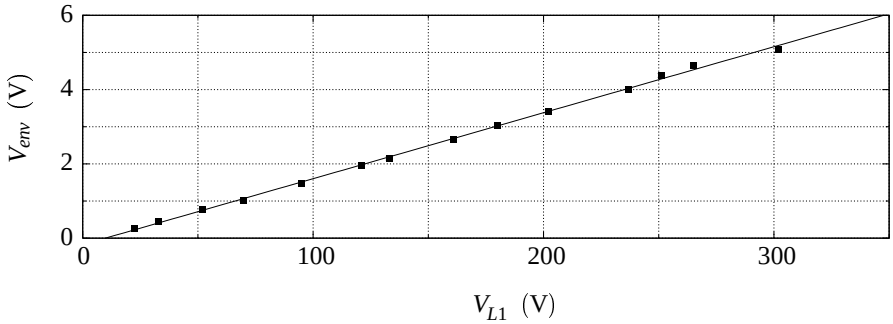


Figure 2.16: *Test results for the amplitude detector on the transmitter board. The frequency was 2.0 MHz, the supply voltage 6 V, and the coil inductance 7.0 µH.*

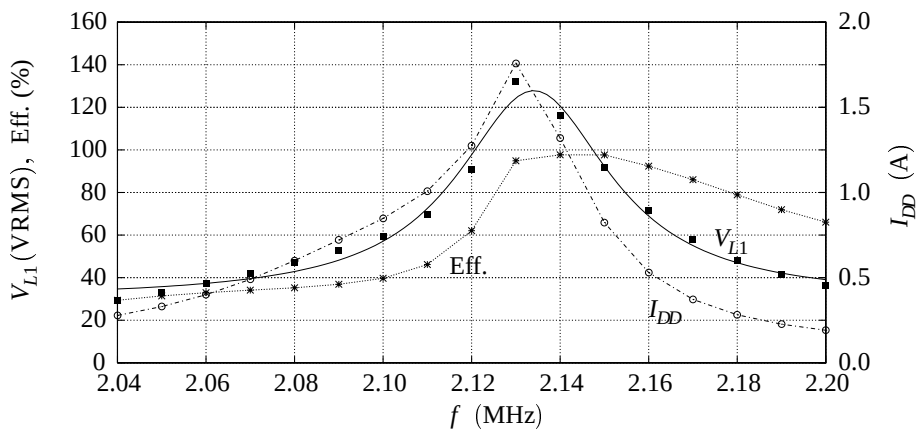


Figure 2.17: *Results for a transmitter frequency sweep with an LC load consisting of the 2.65 µH coil in figure 2.14. and a tuning capacitance of 2060 pF. The supply voltage was 6 V. The Q of the LC tank is approximately 49.*

counterparts would need. They could have been replaced to reduce the current overhead in the transmitter.

An interesting feature of the efficiency graphs is that the efficiency seems to be higher at frequencies just above resonance than just below. This is probably due to an interaction between the tuned load and the switches, where some parasitic effect is reduced when the tank period is slightly shorter than the switch period.

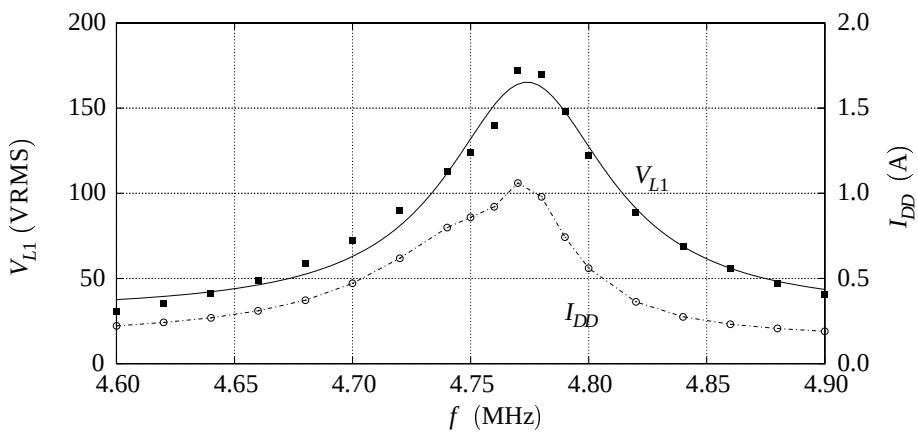


Figure 2.18: *Results for the same setup as in figure 2.17, but with a tunig capacitance of 400pF.*

3
RECEIVERS

3.1 Requirements

Receivers for implanted systems must be designed for conditions which are unusual in several respects. Due to the dual power/data nature of the link, the signal amplitude is far greater than the norm in radio receivers, being equal to or even greater than the supply voltage. Where normal RF receivers have a low-noise amplifier as their first stage, a receiver for an inductive link must often include an attenuator as the first stage. The low-noise considerations which drive much of today's receiver design therefore have little importance in this case. The special efficiency requirements towards the transmitter for an implanted system set some limitations on the range of available modulations schemes, and modulation amplitudes. As we showed in section 2.4, the prevalent transmitter designs can more easily generate amplitude modulation than frequency modulation. For much the same reason, quadrature modulation schemes are not used with such transmitters. Quadrature modulation, or AM-PM schemes in general, have better power efficiency than simple pulse amplitude modulation (PAM) schemes, meaning that they require less transmitter power for the same bit error rate. The bit error rate (BER) of a modulation scheme is usually an important figure of merit, and can along with the spectral efficiency be the most important parameter governing the choice of modulation for an application. The signal power in an inductively coupled system is however so large that comparisons of bit error rates based on the SNR are meaningless. A more meaningful efficiency measure is the spectral efficiency, where n-level PAM is better than n-ary QAM by a factor of 2 [13].

Because the transmitter antennas for these systems are mainly near-field antennas, the transmitted power falls very quickly with distance. Equation (2.21) shows that the field magnitude falls with the third power of the distance in the near field, and the radiated power in the far field is not nearly high enough to affect a receiver. Contrary to many other RF systems where interferers in adjacent channels can have higher amplitudes than the desired signal, the receiver design can ignore phenomena such as intermodulation. Because of the low sensitivity of the receiver, it is unlikely that other electronic equipment can disturb the operation of an inductively coupled implant. That is not to say that it should not be checked against interference from powerful transmitters like mobile phones, airport metal detectors

and some medical equipment like SQUIDs and NMR scanners.

3.1.1 Special conditions

The task of designing a receiver for an inductive link may sound trivial due, based on the preceding description; the SNR is quasi-infinite, the signal power is large and relatively simple pulse amplitude modulation schemes give the best results. There are however a few factors which complicate matters.

Because the carrier is also used to transfer power to the system, there should not be any long periods during which the amplitude is low. When the data link is idle the amplitude should be fixed for maximum power transfer. Such idle periods are quite common, occurring for example between command transmissions to an actuator or between sample points for a sensor. The carrier should therefore not be modulated (at least not PAM) during the idle periods, and the receiver can not use any received data to stay synchronized to the transmitter. A constant-envelope modulation scheme would bypass this problem entirely, but the possibilities were not investigated further because of the incompatibility of frequency shift modulation with our transmitter designs.

Symbol synchronization is not the only function of the receiver which is disturbed by the idle periods. It is impractical to hard-code the detection thresholds in the receiver hardware because this would have a detrimental effect on the versatility and reliability. The receiver must be able to adapt to a range of modulation indexes and carrier frequencies, to be compatible with more than a single transmitter. The detection levels for PAM demodulation must therefore be determined within the receiver. It is in principle possible to store the threshold values between data bursts, but the problem is to know the difference between the unmodulated carrier and a bit sequence resulting in a constant envelope. This problem can be alleviated by including start bits and using suitable encoding schemes as will be described below, but it is still a non-trivial task to make the demodulator respond correctly at the start and end of a pulse.

Many of these problems could be solved by using what amounts to standard operating procedure in today's digital receivers, sampling the analog baseband signal and processing it digitally. Adding a DSP to the chip is however not a trivial task, and the research value of designing a simple DSP is close to zero. Integrating the processor as commercial IP core would not be much simpler. We chose therefore to design the demodulators with fully analog signal processing (with the possible exception of a few digital state elements).

Although the transmitter described in section 2.4 can provide any amplitude modulation index from 0 to 1, more efficient transmitters typically have a narrower range [107]. One requirement towards a receiver is therefore that it should function correctly with small amplitude modulation depths. As a benchmark number we have used a modulation depth of 10%, or a modulation index $m = 0.05$. One can also argue that a small modulation index is desirable to maintain some power transfer at all times. In practice however, the power conversion circuit is so sensitive to amplitude variations that a reduction of only a few percent reduces the transferred power to zero. Power conversion circuits are discussed in detail in section 4.2.

The power consumption of the receiver should be low, just as the rest of any implanted chip should preferably use no power at all. Any reduction in the power consumption of the implant can extend the life of the battery in an external transmitter, so low power is one of

the design goals for the receivers.

The necessary data rates depend on the application, and can range from a few hundred bits per second to megabits per second in a retinal prosthesis [22, 21]. Receivers for simple nerve stimulator or transceiver chips do not need as high bit rates as a retinal prosthesis, so the receivers presented in the following sections are designed for peak bit rates of 100-200kbit/s. The requirements depend on the intended mode of operation, whether the transmission is in bursts or continuous, whether the implant is programmable or needs a command for each action, and so on.

Implanted integrated circuits usually include some digital logic which must be clocked at a stable frequency. The stable frequency may be necessary for stimulation pulse timing [46] or to set the sampling intervals in a sensor system. Unless a crystal oscillator is added to the system, the carrier frequency is the most stable timing signal available. One task of the receiver is therefore also to extract the carrier signal and generate the system clock from it. This may be as simple as to convert the analog carrier to the digital domain with a hard limiter. Some modulation types, like on-off keying PAM or FSK, do not provide a carrier signal which is present at all times with a fixed frequency. In that case, a PLL must be included to provide the system clock.

3.1.2 Frequency allocation

The allocation of frequencies for specific uses varies from country to country, but there are a few common ranges which can be used for biomedical equipment. There are for example many frequency blocks in a wide range around 200 MHz which can be used under the general heading “biomedical telemetry”. These are however not appropriate for the type of radio transmission discussed here, being more useful for active transmitters used in remote monitoring of biological functions.

Lower frequencies are more attractive in connection with inductive transmission for several reasons, practical and regulatory. The relatively low self-resonant frequencies of the large transmitter coils place the upper limit of the frequency range somewhere under 100 MHz. Other practical matters, like the increased absorption in tissue at higher frequencies, lower transmitter efficiency and the reduced power converter efficiency, also push the frequency selection towards lower values. Because of the strict regulation of the RF spectrum, it is not possible to choose operating frequencies at will, but there is a transmitter power limit under which these regulations can be ignored. Figure 2.4 shows that the radiated power depends very much on the frequency; a low operating frequency can help a transmitter to stay below the legal limit.

The radio equipment designed in the course of this project was related more to research equipment than product development. We were therefore not overly concerned with laws and regulations regarding radio frequencies. An awareness of these issues can however help in selecting appropriate operating parameters, and to avoid problems later.

There exist frequency bands in the low MHz range which are reserved, among other things, for low-power radio equipment with loop antennas. These are the *Industrial, Scientific and Medical (ISM)* bands, which exist out to GHz frequencies. The Bluetooth protocol is an example of use for the 2.4GHz ISM band. The ISM bands are not exactly the same in all jurisdictions, there being minor variations in their specification and some of them are not implemented everywhere. A cursory examination of the regulations in force in Denmark,

Center (MHz)	Bandwidth (kHz)
6.78	30
13.56	14
27.12	326
40.68	40

Table 3.1: *Low-frequency ISM band allocation.*

the United Kingdom and the USA reveals the ISM bands shown in table 3.1, which might be used for inductive links. Of these, the 13.56 MHz band is probably too narrow to be useful, and the 40.68 MHz frequency is rather high. The 6.78 MHz frequency has been used for reference in some of the following designs, where a specific frequency is needed.

Because some of the ISM bands are very narrow and the desired bit rates for complex implants can be quite high, other modulation schemes with better bandwidth efficiency than two-level PAM can be considered. We mentioned before that n-level PAM has a better bandwidth efficiency than any other common scheme, so it is probably the one to use. Multilevel PAM does however require accurate detection circuits which can discriminate between the different levels. The use of DSP methods is probably mandatory in that case.

3.1.3 Previous designs

To put things in perspective, we provide here a short overview of recent receiver designs by others for inductively coupled electronic implants. Despite these having been presented during the last 3-4 years, a direct comparison of their power consumption and performance would not be fair since they are implemented in different technologies and have different applications. They can however provide an idea of common design practices and operating parameters for such systems.

Figures 3.1 and 3.2 show samples of demodulator for inductive links. The approaches are very different, but a few common characteristics can still be seen. All use amplitude modulation, and the data is encoded either directly in the ASK (amplitude shift keying) levels or by pulse width modulation of the ASK waveform.

Figure 3.1(a) shows an ASK demodulator for a retinal prosthesis [58]. It is designed for a carrier frequency of 13.56 MHz(an ISM band), approximately 50% modulation depth and Manchester encoding, with a data rate up to 200kbit/s. The power dissipation of the demodulator itself is not specified, but the total stimulator has an internal power dissipation of 5 mW with a 10 V supply. It is implemented in a 1.2 µm CMOS technology. It is based on two-stage filtering of the carrier envelope, and comparison of the outputs of the two passive filter sections which have different time constants. While this circuit reportedly functions quite well, it requires a rather large modulation index and continuous modulation of the carrier.

Figure 3.1(b) shows a similar solution from [122] which is implemented with a full bridge rectifier and a single-ended threshold detector. This demodulator is designed for an implantable microstimulator, with a 2 MHz carrier, 20–30% ASK modulation depth and a bit rate up to 40kbit/s. The power dissipation of the demodulator is not specified, but the total chip uses about 45–55 mW from a 10–21 V supply. The passive filter section requires large component values to implement long time constants, which can be problematic in some

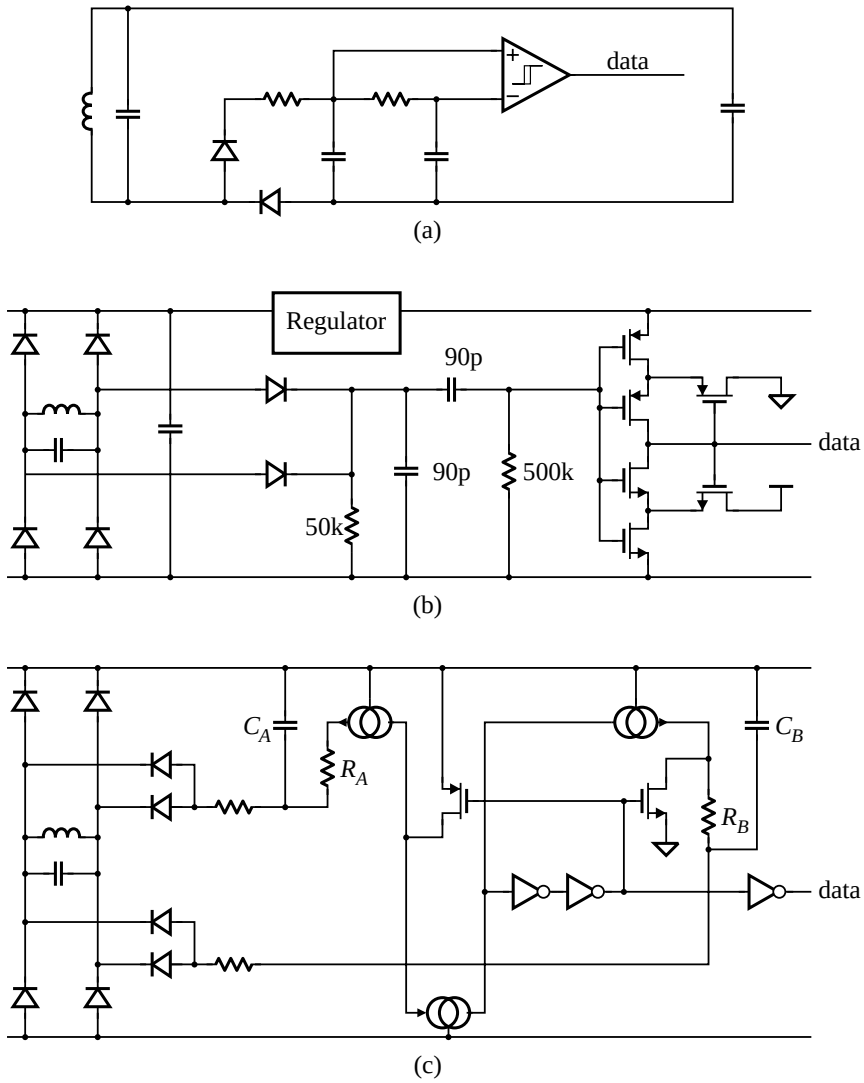


Figure 3.1: *Previous receiver/demodulator designs for inductive transmission. The power conversion circuits in each case are also shown since they have a large influence on the receiver design. Part (a) shows an ASK demodulator for 13.56 MHz by Kolnsberg [58], part (b) shows a 2 MHz PWM ASK demodulator by Ziaie [122] and part (c) shows a 4 MHz PWM ASK demodulator by Akin [4].*

processes. The Schmitt trigger circuit is a rather inefficient type which draws large currents (up to several mA) close to the trigger point. The DC value of the Schmitt trigger input is obviously 0 V, and the circuit is triggered by positive transients of the carrier envelope. Reliance on specific trigger threshold and amplitude values can limit the applicability of the

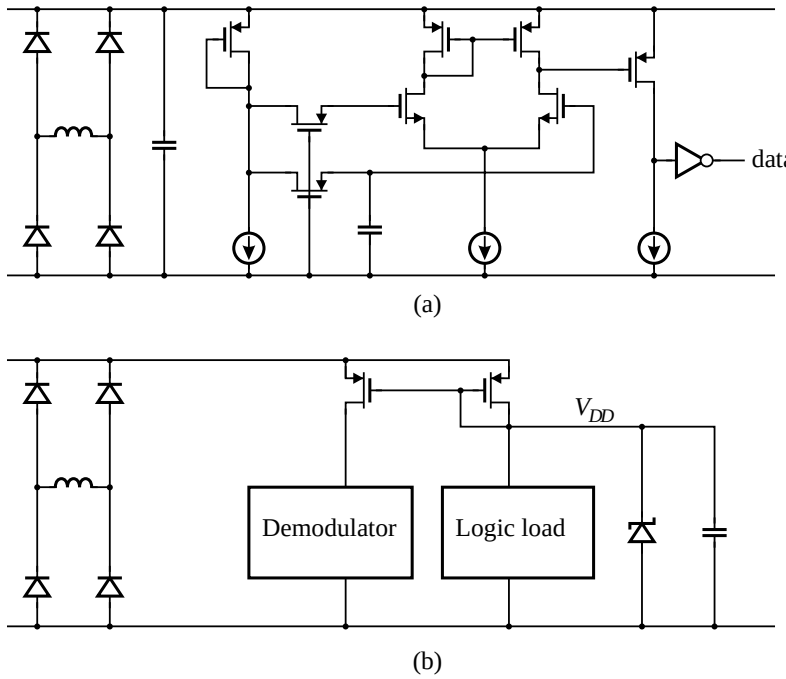


Figure 3.2: *More receiver designs for inductive transmission. Part (a) is from a 5 MHz ASK smart card receiver by Bouvier [17] and part (b) is a 13.56 MHz ASK demodulator by Nebel [70]. The Zener diode in part (b) represents an active shunt regulator.*

circuit.

The circuit in figure 3.1(c) uses two different time constants $R_A C_A$ and $R_B C_B$ to implement a similar high-pass mechanism, but the Schmitt trigger is replaced by a differential current-mode detection circuit. The two MOS transistors shown in the diagram provide hysteresis for the detector. The circuit is from a neural signal transducer for a micromachined sieve electrode [4], and is implemented in a 3 µm BiCMOS process. The demodulator dissipates 2 mW from a 5 V supply. While is does not have the same trigger threshold problem as the previous circuit, it still requires constant modulation of the carrier to set the DC level on the slow RC term.

Figure 3.2(a) shows a demodulator from a CMOS smart card communication interface [17]. There is no explicit input to the demodulator. The carrier envelope is instead sensed through the supply terminal, since there is no overvoltage protection in the system. An overvoltage shunt protection circuit with a sharp knee in the *I–V* characteristics should be included in an implanted IC because of the large power transfer variations which can occur. The conditions in a smart card reader are better controlled because of the fixed geometry. This circuit is also based on a comparison of the same signal filtered through two time constants.

The last example in figure 3.2(b) is radically different from the others, as it uses a supply current sensor to measure the carrier amplitude. The main part of the IC (marked “Logic

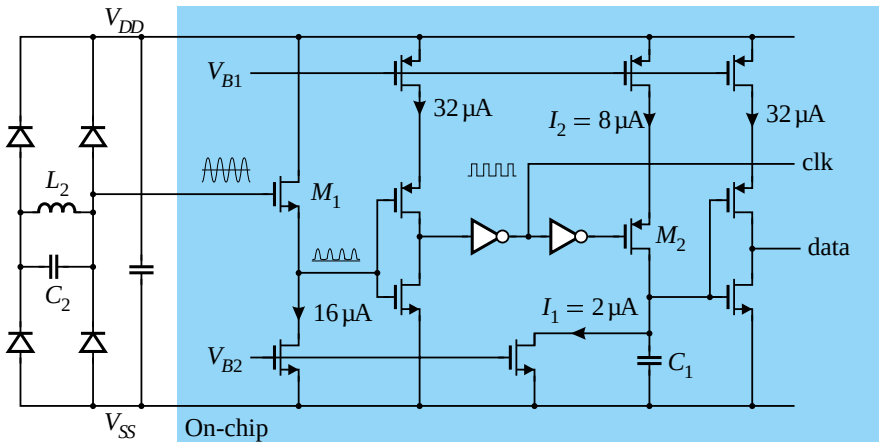


Figure 3.3: *A diagram of the simple receiver showing the on-chip part (shaded) and the external LC antenna and power converter.*

load" since it consists mainly of CMOS logic) receives its supply current through a diode-connected transistor. This current is the mirrored with a ratio of 1/500 to the demodulator circuit, which performs a current-to-voltage conversion, filtering and detection. If this circuit is to function correctly, the transmitter must constantly transmit enough power to forward-bias the bridge rectifier. It is therefore not possible to send data during "free-running" periods, where the chip is using energy available from a storage capacitor. This can be a disadvantage if the device is powered intermittently. The method also introduces a gate-source voltage in series with the supply, reducing the efficiency and supply headroom.

3.2 Basic single-ended receiver

As a way of introduction to more advanced receiver designs, we present here a receiver whose preliminary design was done in a previous project [46], but which was first implemented and tested on the stimulator chip discussed in section 6.1.

This receiver was implemented in an ancient 2.4 µm CMOS technology because of the supply voltage requirements of the system in which it was used. The carrier frequency was fixed at 5 MHz and the data link was run in burst mode, with unmodulated intervals between data sequences. The modulation was simple on-off keying PAM, that is logical '1' was represented by the full carrier amplitude and the carrier was turned off for logical '0'. To aid in the task of symbol synchronization, the modulation was synchronized with the carrier, with each data bit using exactly 50 carrier periods. The data link was used to transmit 32-bit command words to the device and each command was preceded by two start bits to aid in the detection of each command.

A schematic diagram of the demodulator is shown in figure 3.3. The figure does not show the digital state machine which was used to perform the symbol synchronization or the PLL whose task it was to track the carrier and generate the system clock. The signal is taken directly from one terminal of the receiver coil to the gate of the source follower

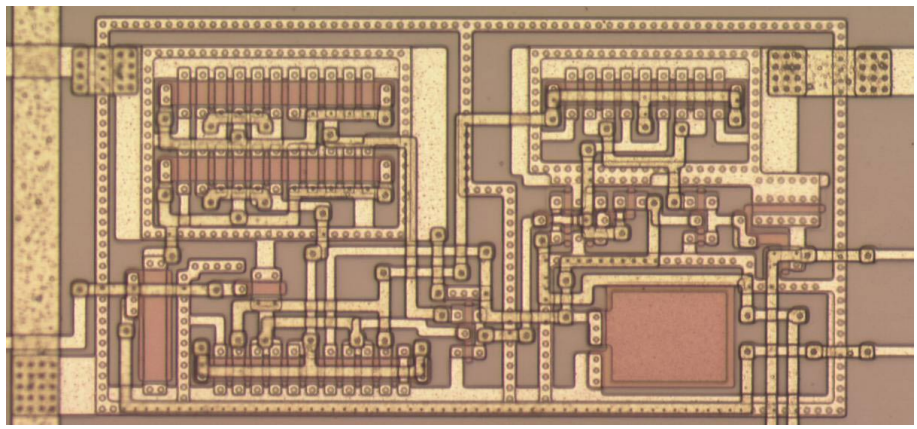


Figure 3.4: *A basic receiver used in a stimulator chip described in [47].*

M1. The output of the source follower is fed to a current-starved inverter, and subsequently to two normal CMOS inverters. This is a basic carrier extraction circuit which generates a square-wave signal with the same frequency as the carrier and a duty cycle slightly less than 50% at the output marked 'clk'. The carrier envelope is extracted by use of the capacitor C_1, which is drained by the current source I_1 and charged through the switch M2 by the current I_2. The relation between the current amplitudes, the switch duty cycle (which is equal to the duty cycle of the 'clk' signal) and the capacitance determine the response of the capacitor voltage. This capacitor voltage is then processed by yet another current-starved inverter to produce the 'data' signal, which should represent the carrier envelope.

Testing of the chip with this receiver circuit showed that it functioned as expected, but it does however suffer from a few problems. The fixed bit/carrier period ratio is a disadvantage for two reasons. In the first place, our experience showed that building a synchronous transmitter was rather tedious, and forced us to accept a few compromises in its design. In the second place, the on-chip PLL had a tendency to lose a few cycles during the '0' periods in each command word transmission, which led to a fair number of transmission errors.

The direct connection to the gate of M1 was also a potential ESD problem. Input protection diodes were included on the chip, but these are connected in parallel with the external power converter diodes and can, if the relative diode voltage drops are shifted in the wrong direction, lead to large current transients through the protection diodes and eventually to their destruction.

The receiver requires full on-off keying, that is a modulation depth of 100%. As we mentioned before, this type of modulation cannot be provided by all transmitter types. The test transmitter in figure 2.13 could be used for on-off keying, but that transmitter is unusual and suffers from a less than optimal efficiency.

As a last complaint against this design, it should be mentioned that the design did not take into account one important effect. When the transmitter is turned off to transmit a logical '0', the common mode voltage of the antenna is essentially floating, and the capacitance on the input node must be discharged to produce a '0' at the source of M1. This was implemented by adding a small conductance between the input node and ground, and could have been

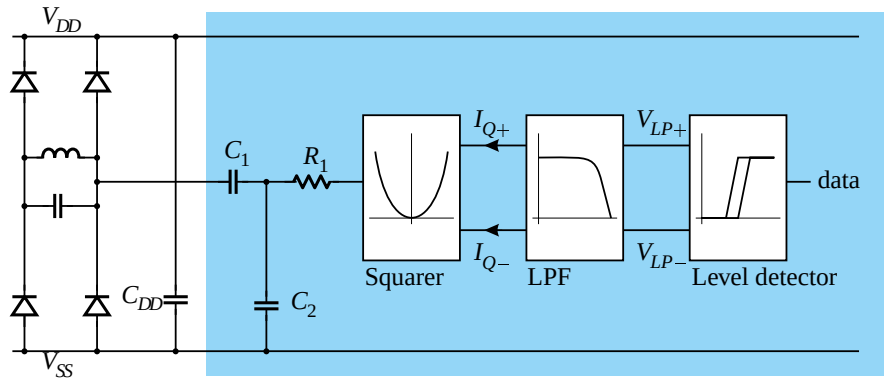


Figure 3.5: *A block diagram of the demodulator circuit.*

done on-chip.

Overall, this receiver design was deemed unsuitable for further use, and other solutions were sought.

3.3 Low-power ASK demodulator

To obtain an improvement over the previous design, a new receiver was designed as a separate project. The goal with this design was to create a low-power receiver/demodulator which could be used with small modulation indexes, and was free of problems related to ESD or capacitive parasitics at the off-chip interface. In addition, we wanted a receiver with an adaptive detection level so that it need not be hard-coded into the design, and which could handle the start of burst transmissions in a sensible way. The carrier frequency was to be free in the range 1–10 MHz, covering the normal range of inductive links. As a target, the bit rate was set at 100kbit/s, with no specific relation between the carrier and bit periods.

A block diagram of the demodulator is shown in figure 3.5. It uses the conventional tuned LC antenna, with the power for the system obtained from the RF carrier. A power conversion circuit was not implemented on this chip, but added externally with discrete diodes for testing. The capacitive divider performs the dual task of protecting the IC against the large transients at the antenna and attenuation of the signal. The capacitors are implemented in metal layers with thick a oxide dielectric, so their breakdown voltage is a few hundred volts. The output of the divider is connected to a current-mode squarer circuit whose task is to produce both an extracted carrier signal replica and the signal envelope. The envelope data is then filtered in an active third-order lowpass filter. The output of the LPF is connected to a new level detector which extracts the digital modulation data from the envelope information.

3.3.1 Passive input network

The use of thick dielectric metal capacitors provides enough protection for the internal circuitry against external transients. They are however not enough to protect the internal circuit against ESD, as such voltage transients are frequently on the order of kilovolts. ESD diodes

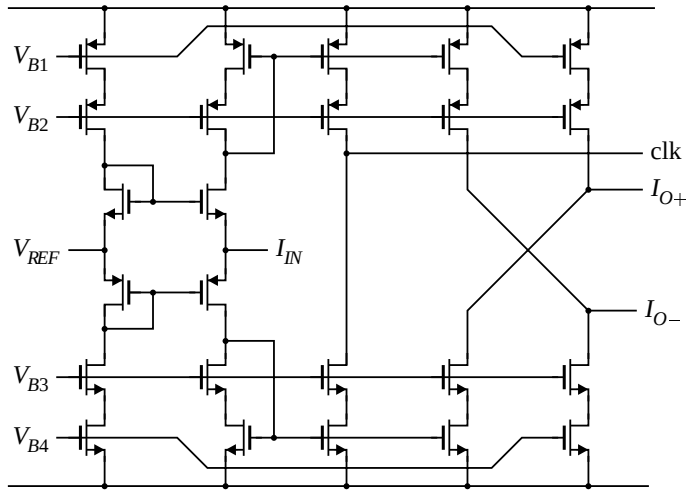


Figure 3.6: *The current-mode squarer circuit. The bias current in each branch is 1 µA.*

were therefore added on the internal capacitor terminals to provide some protection for the squarer input. These diodes are always reverse-biased during normal operation, in contrast to the diodes in the previous receiver, and their only effect is to add some capacitance.

The output impedance of the tuned antenna circuit is on the order of 1 kΩ and we can consider it as a voltage source as long as the load impedance is much larger. The carrier frequency is close to the resonant frequency of the tuned circuit so we may assume that the output impedance is resistive. This source is connected to the capacitive step-down transformer consisting of the input capacitors, R1 and the input resistance of the squarer. The network consisting of the source resistance, the capacitors and the load resistance has a bandpass transfer function. As long as the lower and upper corner frequencies are well separated (which they are) the transfer function in the middle of the band is approximately

$$\frac{I_{IN}}{V_A} = \frac{C_1}{(C_1 + C_2)(R_1 + R_i)} \tag{3.1}$$

The lower and upper corner frequencies are $1/\omega_l = (R_1 + R_i)(C_1 + C_2)$ and $1/\omega_u = R_S C_1$ respectively

3.3.2 Squarer

The fundamental block in the envelope detection mechanism is the current-mode squarer shown in figure 3.6. This circuit is similar to a class AB CCII+ current conveyor [19] but the connections of the current mirrors and current sources implement a square-law transfer characteristic. This squarer is derived from a current-mode multiplier published in [76], which uses the square-law characteristic of MOS transistors in strong inversion. The differential output current is given by

$$I_{O+} - I_{O-} = \frac{I_{IN}^2}{8I_B} \tag{3.2}$$

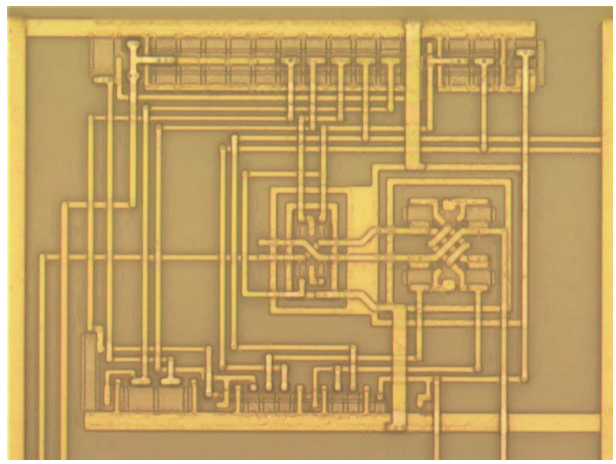


Figure 3.7: *A die detail showing the squarer circuit. The squarer is approximately 100 µm on a side.*

where I_B is the bias current of the input transistors. This behavior is derived by assuming that the p-channel and n-channel transistors in the input quad are matched. This matching can obviously never be performed because of the independent process variations of the two transistor types. While this mismatch would be unacceptable in a precision circuit it does not have a severe effect on the envelope detection in the demodulator.

One of the features which makes this circuit attractive for our purposes is that a relatively high small-signal bandwidth can be achieved with low power levels. With the nominal bias current of 1 µA in each branch, or a total current of 5 µA, the 3-dB bandwidth hinted at by simulations is greater than 30 MHz. The supply current of the squarer is actually signal-dependent due to the class AB operation, and the time average rises from 5 µA to 10 µA at the maximum specified signal amplitude.

The carrier signal can be transformed into the digital domain for use as a clock or time reference by a simple and inexpensive addition to this circuit. The two current mirrors used in the squaring function were given secondary current outputs, and these were connected together. This high-impedance node slews to the positive supply or to ground according to the sign of the input current, thus creating a clock signal corresponding to the carrier frequency.

Other CMOS multipliers/squarer were considered for this use. There is for example one squarer circuit presented in [65] which could have been used. It does however have a differential voltage-mode input which would have required a different input circuit, and it uses more current than our solution for equal performance. Other CMOS multiplier circuits are not as easily adapted to operate as a squarer as they require preprocessing of the input voltages by forming sum and difference terms [30, 28].

3.3.3 Low-pass filter

The low-pass filter used to extract the envelope data from the squared input signal is shown in figure 3.8. It is a differential third-order filter, consisting of a passive first-order section

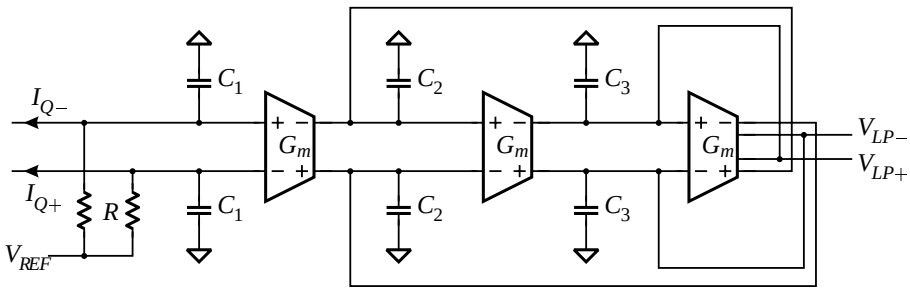


Figure 3.8: *The differential LPF implementation.*

and an active section which is implemented as a G_m–C filter. Many of the same rules apply to very low-power G_m–C filters as to high-speed ones. Both are limited in performance by parasitic capacitances. In our case it is the speed/current tradeoff rather than the noise/current tradeoff which is most important. As is pointed out in [69], it is therefore important to use a filter topology in which all signal nodes have a desired capacitance to ground. Furthermore, the transconductors used in the filter may not have any internal high-impedance nodes. These requirements are fulfilled by the topology shown in figure 3.8 and by the transconductors in figure 3.9.

Identical transconductors are used in the filter, for simplicity and better matching of the transfer function. The filter has one pole at $\omega_1 = 1/RC_1$ due to the passive first section. The second-order active section has a pole frequency of

$$\omega_2 = \sqrt{\frac{G_m^2}{C_2 C_3}} \tag{3.3}$$

and a damping coefficient given by

$$\zeta = \sqrt{\frac{C_2}{4C_3}} \tag{3.4}$$

An approximation of a third-order Butterworth transfer function is obtained by selecting the component values so $\omega_1 = \omega_2$ and placing the poles at 60-degree intervals.

The cutoff frequency of the filter is set with respect to the highest expected modulation frequency, with a generous margin for process variations. The nominal frequency is 350 kHz, to permit unencoded data rates of up to approximately 200kbit/s.

Figure 3.9 shows the transconductor used in the filter implementation, along with the common-mode feedback circuit. This CMFB circuit is necessary at the two internal nodes of the active section. It is a current-steering type which causes relatively little distortion of the differential mode signal [32]. The transconductor itself is based on a simple differential pair, with no special precautions to eliminate the relatively large distortion. The nonlinearity of the transconductor is of little importance at the applied signal level (up to 200 mV differential). The 3dB-frequency of the transconductors is approximately 10 MHz, which is sufficiently high not to affect the transfer function of the filter.

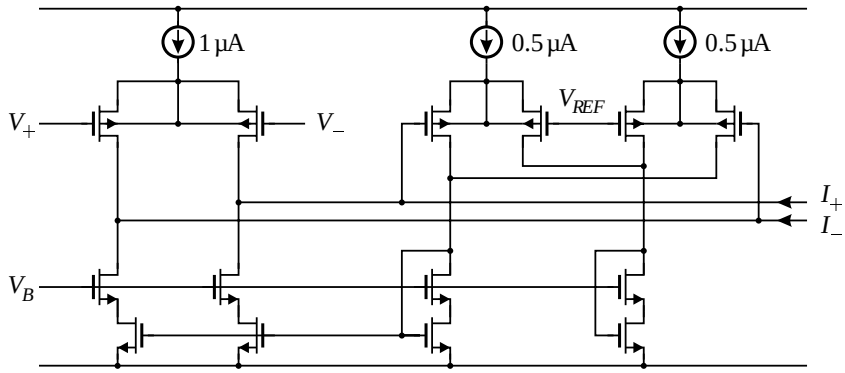


Figure 3.9: *A differential transconductor for the low-pass filter, with common-mode feedback circuit.*

3.3.4 Level detector

The level detector used in the receiver is specially adapted to use with burst data transmission where the modulation levels are not fixed in the hardware. It can adapt to different modulation levels, and requires only two start bits to sense the modulation levels. The absence of modulation in the intervals between data bursts make it impossible to use a simple LP filter with a long time constant to store the average signal level. Since we chose not to use a complex and power-hungry DSP solution, this adaptive level detector was developed. The start bits are used to provide information about the high and low modulation levels, but if a biphase encoding scheme like Manchester coding is used, a single start bit is sufficient.

A diagram of the detector is shown in figure 3.10. It uses three single-ended transconductors with multiple outputs. All three transconductors have the same transconductance G, except for the third output of G_1 which has twice the magnitude of the other outputs, $I_3 = 2I_2 = 2I_1$.

The positive peak voltage at the input is stored on C_p (referred to V_{REF}), while the negative peak is stored on C_N. In the following calculation of the transfer function from the input to the positive peak voltage V_P, the diode transistor M_1 can be ignored. The justification for this will be shown below.

From the diagram, we have the small-signal currents $i_1 = G(v_{LP-} - v_{LP+}) := -Gv_d$ and $i_p = G(v_P - v_{REF}) = Gv_P$. The voltage on the capacitor is $v_P = -(i_1 + i_P)/sC_P$. Inserting the expression for the currents, and solving for the transfer function, we get

$$\frac{v_P}{v_d} = \frac{1}{1 + sC_P/G} \tag{3.5}$$

Two observations can be made at this point. In the first place, the DC value of this function is 1 (this result presumes perfect matching of the transconductances), so a true copy of the peak voltage is effectively stored on C_P. In the second place, this is a first-order LP function, so the voltage on the capacitor approaches the final value with no overshoot. This last fact is the justification for omitting the diode-connected transistor M1 in the analysis. The result for the negative peak voltage is similar. The time constants C_P/G and C_N/G should be set

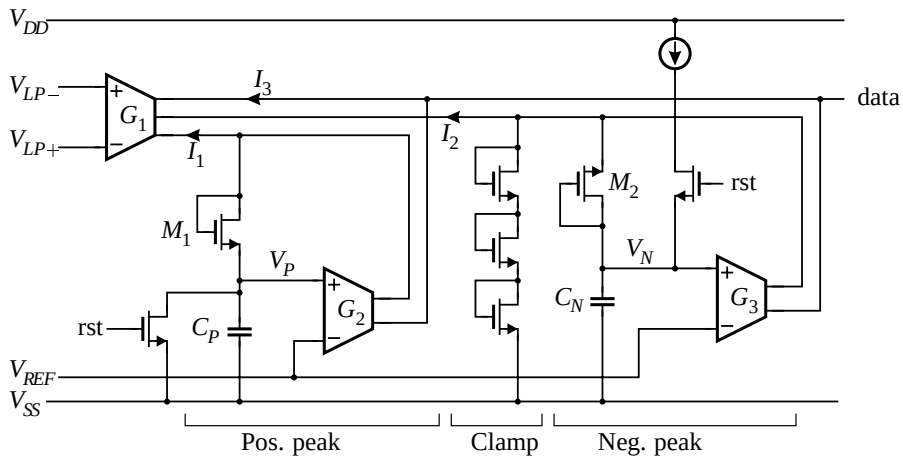


Figure 3.10: *The peak detector and data signal extraction circuit.*

in relation to the length of the start symbol, to allow the positive and negative peak voltages time to settle.

The digital 'data' signal is extracted from the input signal by comparing it to the average of the two peak values. This is done by summing the currents $I_3 = -2GV_d$, $I_P = G(V_P - V_{REF})$ and $I_N = G(V_N - V_{REF})$ at the output node, and letting it slew to the positive supply or to ground according to the sign of the result.

The reset signal is used to initialize the values of the capacitors which store the positive and negative peak signal voltages. The positive peak voltage V_P is set to V_{SS} by the reset signal, while the negative peak voltage V_N is set to approximately four gate-source voltages. This last value is determined partly by a voltage clamp consisting of three diode-connected transistors. The purpose of the clamp is to limit the voltage swing on the summing node for the negative peak computation. This reduces the amplitude of the capacitive feed-through to C_N through the diode. No clamp is necessary for the positive peak since the available voltage range at the summing node is more limited. The negative peak reset transistor is not connected to V_{DD}, but to a current source, in order to limit the current through M2 and the clamp during reset.

3.3.5 Experimental results

The demodulator was implemented in a standard digital 0.5 µm CMOS process, with one polysilicon and three metal layers. The maximum supply voltage is 3.3 V.

All measurement results were obtained with a supply voltage of 3 V. In order to test the circuit with a wide range of carrier frequencies, modulation parameters and signal levels, a signal generator was used for most measurements. The functionality of the circuit was however verified with transmission through an inductive link with a carrier frequency of 5 MHz at 200kbit/s.

Due to the nature of the dual power/data link, the peak-to-peak amplitude of the unmodulated signal is almost constant, and equal to the supply voltage plus two diode voltage drops.

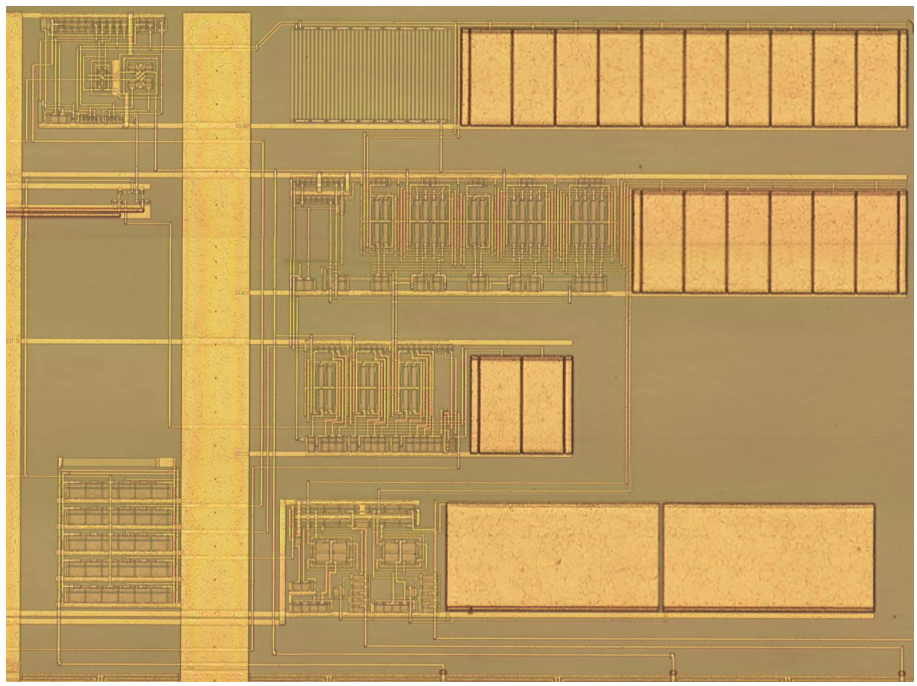


Figure 3.11: *A photo of the demodulator core. The figure shows the squarer in the upper left corner, the passive LP section in the top row on the right, the active G_m–C filter in the second row, the adaptive level detector in the third and some buffers and bias in the last row.*

In the following test results, the unmodulated signal level is fixed at $3.5\,V_{p\text{-}p}$.

The sensitivity of the carrier extraction mechanism is limited by the parasitic capacitance on the 'clk' signal output node in the squarer (figure 3.6). The necessary signal level for detection rises as the frequency increases, since the current into the output node is proportional to the signal. An output buffer was connected to the 'clk' signal, to drive the signal off-chip for testing purposes. The parasitic capacitance on the signal node was higher than necessary, and the sensitivity of the carrier extraction does therefore fall off faster with frequency than expected.

The measurements show that the detection threshold is approximately $0.4\,V_{p\text{-}p}$ up to 5 MHz, and rises to $3.6\,V_{p\text{-}p}$ at 12 MHz. Simulations of the circuit without the added parasitic capacitance show a threshold of $0.8\,V_{p\text{-}p}$ at 15 MHz.

The circuit works as expected for carrier frequencies between 1 and 15 MHz, with 10% AM modulation rates up to 200kbit/s. Figure 3.12 show the results for a 10 MHz carrier modulated at the maximum bit rate.

The supply current of the demodulator was measured for a wide range of operating conditions. At the nominal input signal level of $3.5\,V_{p\text{-}p}$ the supply current was constant at $20.0 \pm 0.5\,\mu A$ over the carrier frequency range of 1–15 MHz. A weak dependence on the signal amplitude was found, with the supply current decreasing by a few percent as the amplitude was reduced to zero. This dependence is due to the class AB operation of the squarer

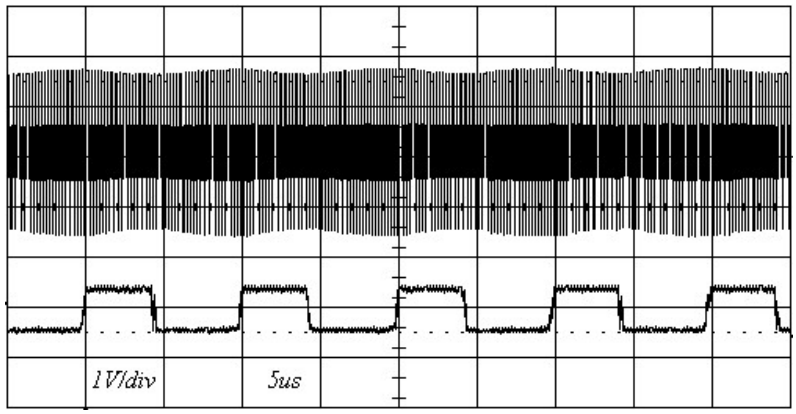


Figure 3.12: *Measurements for a 10MHz carrier with 10% sinusoidal AM modulation. The data output of the demodulator follows the AM envelope correctly, and shows that the modulation level detection mechanism functions as expected.*

circuit.

4

POWER SUPPLY MANAGEMENT

The trend towards constantly increasing system integration is evident in the field of implantable systems as it is in other fields of electronic design. This trend does complicate the design of integrated sensor/actuator and processing like the typical biomedical implant IC. In addition to the difficulties arising from the diverging technology requirements of analog and digital integrated circuits, mixed-signal designs are characteristically plagued by crosstalk between the two domains. The supply disturbances generated by noisy digital circuits can easily couple into sensitive analog circuits. This problem can be tackled in several different ways; for example by reducing the noise generated by the digital circuits, reducing the susceptibility of analog parts to noise and reducing the coupling between the two.

In order to effectively apply any crosstalk reduction strategy it is necessary to have a detailed knowledge of the characteristics of the power supply path from the external power supply into the chip and to the different circuit blocks. A step in that direction is to analyze the influence of the elements in the power supply path on the supply quality.

The DC power supply voltage in an inductively powered implanted system is derived from the RF carrier by rectification and filtering. These two tasks can be carried out on-chip to a varying degree, with complete integration of the power supply conditioning circuits representing the ultimate design goal. Complete integration is attractive for the usual reasons of reliability, minimized interconnect complexity and reduced physical size.

An important limitation imposed by the complete integration of the supply in implanted systems is the compromised isolation between the analog and digital parts. While the inclusion of digital circuitry on analog signal processing ICs has become the rule rather than the exception, the availability of separate filtered off-chip supplies for the two signal domains helps to isolate the analog blocks from the noise generated in the digital part. In the completely integrated case there is only one common supply voltage which must be shared between the two, and it matters little whether this supply voltage is regulated on-chip or off. Basic precautions like using separate package pins for the two supplies, “quiet” ground guard rings and similar techniques [75] are nearly useless since the supply nodes are ultimately connected to a common point.

The demands upon the quality of the power supply can be extremely strict in systems which include nerve signal amplifiers [77, 72] because of the small amplitude of the signals, which is typically on the order of $\pm 10\mu V$ in a low-frequency band from DC up to a

few kHz [104]. Even though the worst noise offenders in the power supply path, the digital logic blocks and the power converter, concentrate their output in the low MHz range and at harmonics of those frequencies, downconversion by nonlinearities and saturation or desensitization of the amplifier inputs can destroy the integrity of the signal. These effects can be mitigated to some extent by using differential signal processing and other noise suppression techniques. A typical low-power amplifier using very large transistors in weak inversion biased with relatively low currents [72] does however suffer from low PSRR at high frequencies because of the strong capacitive feed-through paths existing in such a design. Inevitable mismatches between the differential branches will lead to common-mode to differential conversion of the supply noise into the signal path. Out-of-band supply noise must therefore be reduced as far as possible to allow the use of advanced low-power amplifiers in an implanted sensor.

Some of the generic processing elements which can be identified in most implanted systems powered by inductive links are the rectifier (or power converter) which converts the RF carrier into a useful DC supply voltage, an energy storage element, an overvoltage protection circuits and voltage regulators for different on-chip supply levels. We shall present here some circuits which have been developed for these purposes and are characterized in part by the fact that they are implemented in a pure CMOS technology. Previous solutions have relied either on more advanced processing or on external components. With the exception of the energy storage element, which cannot be placed on-chip by force of energy density limitations (except in systems like smart cards which have either very relaxed power supply requirements or are driven by an extremely powerful RF transmitter), we have integrated all the power supply components on-chip.

The circuits in the following sections show the proposed solutions for the necessary power supply components. A supply voltage level of 3 V is used as a reference since the circuits were implemented in a 0.5 µm CMOS technology with a maximum supply voltage of 3.3 V(+10%). Where specific supply voltages are mentioned, as in the experimental results, they therefore refer to a 3 V supply level. These supply components were incorporated on a test chip with a target current consumption of 100–200 µA, and they are scaled for that supply current. With the exception of specific experimental or simulation data, the results can however be scaled to different voltage or current levels.

4.1 Shunt regulator

The strong dependence of the transmitted power through an inductive link on the link distance leads to large power fluctuations with small geometry shifts. Combined with the ease with which CMOS circuits are destroyed by high supply voltages, this makes the inclusion of some overvoltage protection device a necessity. This has traditionally been solved in implants by the use of zener diodes [50] (external or internal, technology permitting) or by simple p-n junction cascades as shunts across the supply [4].

An alternative is to implement the power conversion part of the IC using high-voltage devices and thereby reduce the risk of destruction by high voltage input transients. This requires the use of devices with breakdown voltages on the order of 100 V and step-down regulators to feed the “normal” low-voltage parts of the IC [50, 122, 68]. In plain CMOS however, the available active devices are the two MOS transistor polarities, low-quality diodes

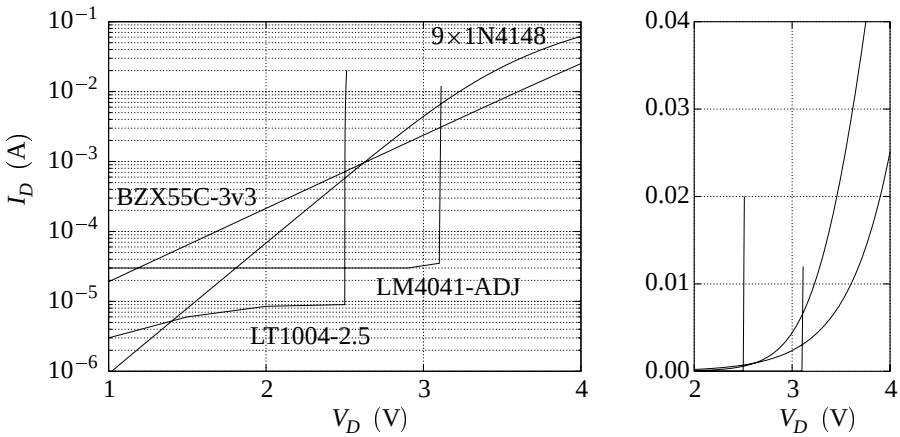


Figure 4.1: *A comparison of some shunt regulator possibilities. The simple diode and zener curves are based on simulation data and the active device curves are taken from the data sheets for the LT1004-2.5 [25] and the LM4041 [26]. The graph on the right shows the same data with a linear current scale.*

and passive devices. Different methods must therefore be employed to obtain comparable performance.

An active solution has several advantages over the passive diode-based one, provided that the performance can be brought up to the same level or better with respect to the critical specifications. A sharp knee in the current-voltage characteristic is one desirable characteristic which allows the system supply voltage to approach the overvoltage protection threshold without activating the shunt device. Likewise, a low small-signal resistance is desirable in the active region to set a predictable supply voltage independently of the shunt current, and to suppress the effect of load and input transients. Zener diode and p-n junction cascades do not fulfill these two criteria since the roughly exponential characteristic does not exhibit a sharp knee, and a lowered output resistance is expensive in terms of current. For a cascade of m p-n junctions for example (with $m \approx 6-9$ for a 3 V supply), the output resistance is $R_{out} = mU_T/I_D$. The zener diode implementation has a similar tradeoff between current and output resistance.

A sharp knee can only be achieved by an active solution, as shown in figure 4.1. The figure compares the two passive shunts mentioned above to two recent commercial micropower shunts. These devices are used for comparison since no similar examples could be found in the research literature. Active shunt regulators are typically used as voltage reference circuits, and as such they are optimized for high initial accuracy, low temperature dependence and low small-signal resistance. The last characteristic, which increases their tolerance towards biasing variations, is what makes them interesting as overvoltage protection devices in an implant. The accuracy and temperature independence are of secondary importance, which is fortunate since those characteristics are typically obtained by trimming and the use of bipolar or BiCMOS processes.

The topology which we have used to implement our shunt regulator consists of a large pass transistor between the supply rails, controlled by an active regulator circuit as shown in

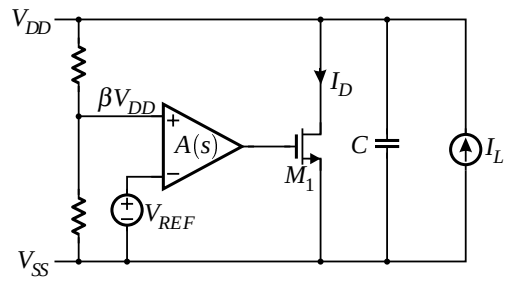


Figure 4.2: *A simplified diagram of a possible shunt configuration.*

figure 4.2. This is the best arrangement if the device is to draw a low current in the off state, and still be able to shunt large currents from the supply in the on state. The choice of pass device is free, subject to what the technology has to offer. In CMOS, it is best to use an n-channel transistor because of the lower gate capacitance per unit transconductance (at least in strong inversion). The lower capacitance reduces the drive requirements for the active device, and therefore reduces the bias current necessary to attain a given transient response.

4.1.1 Regulator stability

The regulator is subjected to load variations from two sources, the input power from the link and variable power used by the system. In an active regulation scheme, the transient response of the regulator must be examined and any misbehavior corrected.

We will now proceed to show that the basic regulator structure in figure 4.2 suffers from severe stability problems. Its transfer function $Z(s)$ with respect to external load current sources has poles that are very close to the imaginary axis. In a first approximation, the amplifier in figure 4.2 is assumed to have the transfer function $A_1(s) = \omega_1/s$, that is a GBW equal to ω_1 and infinite DC gain.

The voltage divider is used to scale the supply voltage down to a suitable level for comparison with the reference voltage. This divider can be implemented using either passive or active elements. The use of a simple resistive divider, while providing a well-defined division ratio, can lead to excessive current drain due to the lack of linear high-resistance layers in many processes. An active divider made up of diode-connected transistors on the other hand, can be dimensioned for a low current drain and limited die area. The tradeoff is that precise divider ratios are impossible, except possibly for low integer ratios. In the following, the divider is assigned a ratio β ($\beta < 1$), which is treated as a small-signal parameter for the purpose of the analysis. The parameter is also assumed to be frequency-independent, an assumption which is valid at the frequencies of interest in a reasonable layout, either using high-resistance polysilicon or MOS transistors. The bias current through the divider is neglected in the stability analysis, as it can be absorbed into the load current.

The boring details of the analysis are left as an exercise for the reader, but straightforward manipulations of the system equations result in the transfer function

$$Z_1(s) = \frac{v_{dd}}{i_L} = \frac{s}{s^2 C + g_m \beta \omega_1} \tag{4.1}$$

The poles are actually on the imaginary axis, implying a neutrally stable system.

A more accurate transfer function for the system is obtained by taking into account the finite DC gain of the amplifier, and setting its gain to

$$A_2(s) = \frac{A_{dc}\omega_0}{s+\omega_0}, \tag{4.2}$$

where $\omega_0 = \omega_1/A_{dc}$ is the 3-dB frequency. This leads to an impedance transfer function for the regulator which can be expressed as

$$Z_2(s) = \frac{s+\omega_0}{s^2C + s\omega_0 C + g_m\beta A_{dc}\omega_0} \tag{4.3}$$

This impedance can rightly be considered to be the power supply source resistance in the system. At low frequencies, the impedance is just the transimpedance of the shunt transistor divided by the loop gain, and at high frequencies the impedance is capacitive with a value controlled by the supply capacitor C. The impedance between these two extremes should preferably be free from peaking.

The pole frequencies in the transfer function (4.3) are

$$\omega_p = \sqrt{\frac{A_{dc}g_{m1}\beta\omega_0}{C}} = \sqrt{\frac{g_{m1}\beta\omega_1}{C}}, \tag{4.4}$$

and the damping coefficient is

$$\zeta = \frac{1}{2A_{dc}}\sqrt{\frac{\omega_1 C}{g_{m1}\beta}} \tag{4.5}$$

As expected, the coefficient is nonzero, although it is very small. Using approximate figures[1] values for a low-power amplifier in a modern technology, and reasonable values for the other elements, results in $\zeta = 0.03$. This can of course be improved by reducing the gain of the amplifier, at the cost of an increased DC resistance. Figure 4.3 shows the influence of amplifier gain on the transfer function.

The shunt bias current provides another control possibility, but it is not really useful for stability control. The g_m of the shunt transistor appears inside the square root, and it therefore needs to be changed twice as much as the amplifier gain (on a logarithmic scale) for the same effect. Increasing the bias current also moves the pole frequencies out towards higher frequencies, but it does not eliminate the tradeoff between DC resistance and peaking. The GBW of the opamp can be increased to improve the damping coefficient, but this is also expensive in terms of supply current, even more so because of the square root.

The solution to these stability problems is to add another form of compensation. The simple regulation method in figure 4.2 is a proportional regulator. Adding the time derivative of the supply voltage into the loop adds a degree of freedom to the regulator design. In control theory terms then, the solution is to use a PD regulator.

The differential term in this case is the time derivative of the supply voltage on the capacitor C, which is as usual the sum of the currents into the capacitor. The three components of this current are the input current from the power link, the supply current of the system and the shunt current through the regulator. The first two of these can be difficult to measure,

[1] $A_{dc} = 70\text{dB}$, $\omega_1 = 2\pi\text{MHz}$, $C = 500\,\text{nF}$, $g_{m1} = 10\mu\text{A}/nU_T$ and $\beta = 1/3$.

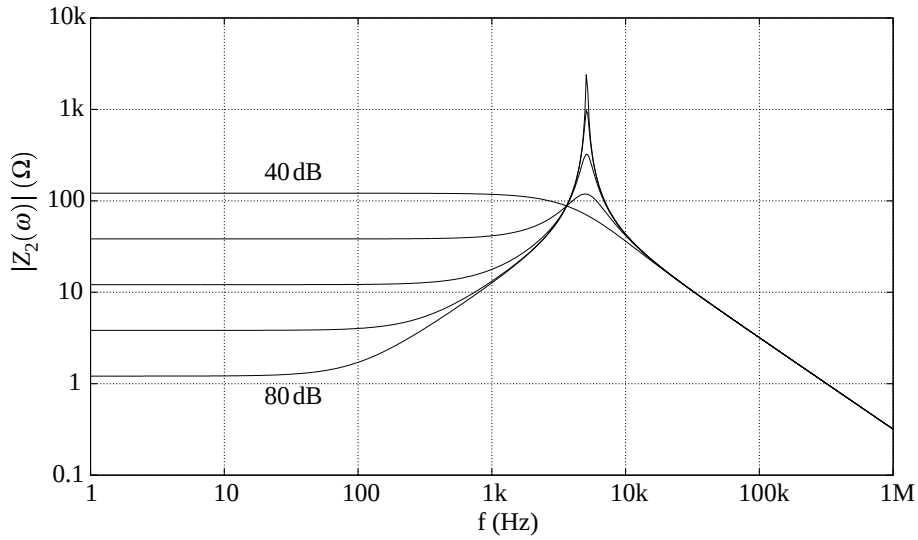


Figure 4.3: *The simple shunt regulator output impedance $|Z_2(s)|$ as a function of frequency, for an amplifier gain of 40, 50, 60, 70 and 80 dB, with $C = 500\,nF$ and $I_{d1} = 10\,\mu A$.*

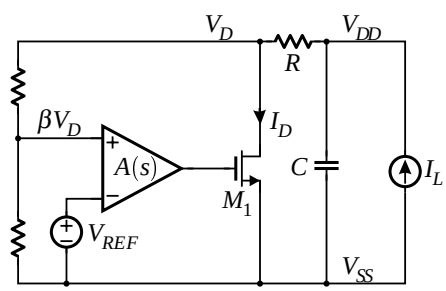


Figure 4.4: *A modified shunt regulator circuit with a compensating resistor.*

but that is not really necessary. By measuring only the shunt current it is possible to reduce the power supply transients due to insufficient stability inside the regulator. The noise contribution from the power link is at such a high frequency that it cannot be eliminated using an active circuit (at least at micropower current levels).

With proportional and differential regulation, the state equation for the system can be expressed as

$$I_{d1} = \alpha(s)\left(\beta V_{dd} - V_R\right) + \gamma(s)\frac{dV_{dd}}{dt} \tag{4.6}$$

The only part of the derivative that needs to be included in the regulation loop is the one stemming from the shunt current, so a simple implementation of (4.6) is shown in figure 4.4. To simplify the notation, we define a new gain transfer function containing the divider factor and the transconductance of M1: $B(s) = \beta g_{m1} A(s)$. The load current is split into $i_L = i_d + sCv_{dd}$ (still neglecting the current through the divider). The drain current in the shunt

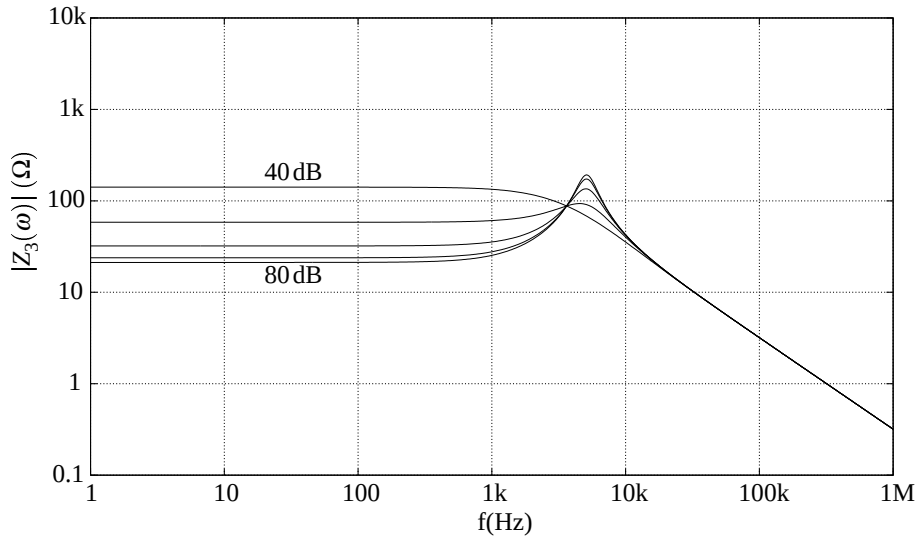


Figure 4.5: *The regulator output impedance with the added compensation resistor $R = 20\,\Omega$. All other parameters are the same as in figure 4.5*

transistor is

$$\begin{aligned} i_d &= B(s)v_{dd} = B(s)(v_{dd} - Ri_d) \\ \Rightarrow \frac{i_d}{v_{dd}} &= \frac{B(s)}{1+RB(s)} \end{aligned} \tag{4.7}$$

After a few more algebraic steps, the output impedance of this third shunt regulator model (in parallel with the capacitor) is

$$Z_3(s) = \frac{v_{dd}}{i_L} = \frac{s+\omega_0(1+Rg_{m1}\beta A_{dc})}{s^2C + sC\omega_0(1+Rg_{m1}\beta A_{dc}) + g_{m1}\beta A_{dc}\omega_0} \tag{4.8}$$

The equation shows that the impedance is the same as for the previous circuit, except for the factor $(1+Rg_{m1}\beta A_{dc})$ which appears in two places. The pole frequency is the same, but the damping coefficient is

$$\zeta = \frac{1+Rg_{m1}\beta A_{dc}}{2A_{dc}}\sqrt{\frac{C\omega_1}{g_{m1}\beta}} \tag{4.9}$$

Figure 4.5 shows the impedance curves parameterized with the gain A_{dc}. The compensation resistor R is not inside the feedback loop, so its value presents a lower limit for the regulator impedance at DC. The figure does indeed show that the output impedance at DC tends towards a lower limit as the gain is increased. There is therefore a point of diminishing returns, beyond which an increase in DC gain only decreases the damping factor.

It may appear possible to use the two free parameters R and A_{dc} to obtain any value of DC impedance and damping, but because of the way in which R enters into the design equations, this is not the case. The operating point of the regulator influences the dynamic behavior of

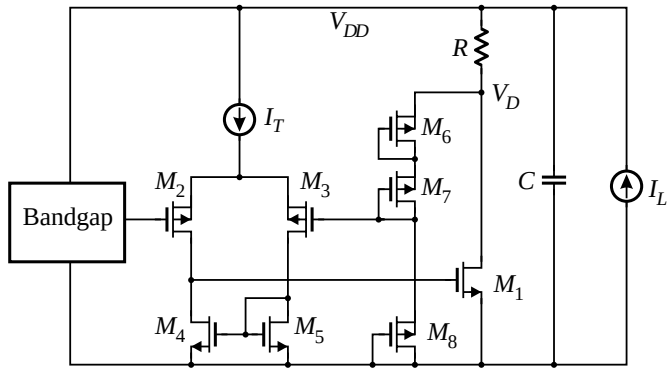


Figure 4.6: *The implementation of the shunt regulator.*

the regulator through the transconductance term in (4.8) and (4.9). We will proceed below to show how a certain minimum performance level may be guaranteed for all operating points by a correct choice of parameters.

4.1.2 Implementation

There are a few other design parameter than R and A_{dc} that can be used to obtain a better damped transfer function than the ones shown in figure 4.5, using the same basic regulator structure. These include the opamp gain-bandwidth ω_1 and to some extent the transconductance g_{m1}. This last parameter is not really a free design parameter, since it depends strongly on the shunt current.

The previous graphs show that a high DC gain in the amplifier is not really useful or necessary. The DC impedance

$$Z_3(0) = \frac{1 + Rg_{m1}\beta A_{dc}}{g_{m1}\beta A_{dc}} \tag{4.10}$$

is equal to a level R_A $(R_A > R)$ when

$$A_{dc} = \frac{1}{\beta g_{m1}(R_A - R)} \tag{4.11}$$

Setting for example $R_A = 2R$ and $R = 10\,\Omega$ and $g_{m1} = 250\,\mu\text{A/V}$ (corresponding to a $10\,\mu$A bias current in weak inversion), we get an amplifier DC gain of 62 dB.

The implementation of the shunt regulator is shown in figure 4.6. A single-stage opamp was used because of the relaxed gain requirements, and because it could fulfill other design goals. The diagram also shows that a bandgap reference was included in the circuit itself. The reasoning behind that is to have a shunt regulator which appears as a true 2-terminal device, and is independent of external biasing and reference source. It also simplifies the testing of the unit. The diagram does not show the connection, but an internal bias signal in the reference was also used to bias the tail current source I_T. The bandgap is a poor version of the reference described in section 5.1. The nominal supply current of the reference is

2.25 µA. The main reasons for the poorer performance are that it is designed for a smaller bias current, and is implemented with a more compact layout. As a consequence of the smaller layout, it suffers a bit more from matching errors.

The voltage divider used in the feedback branch is implemented with p-channel transistors, each in its own floating well. The floating wells were used to eliminate the influence of the bulk effect. With reasonable care in matching, this kind of divider can be used to perform small integer ratio voltage division. However, since the ratio between the desired supply voltage and the reference voltage is not an integer (it is 3.3/1.25, not quite 3/1), it is not possible to use identical transistors. Unmodeled nonlinear effects and process variations can then cause the divider ratio to be somewhat off the mark. This active divider is therefore likely to be the main source of error in the regulator's knee or threshold voltage.

The shunt transistor is dimensioned according to two criteria. The first is that the load capacitance that it presents to the amplifier output determines the dominant pole of the amplifier. The second is that it needs to be capable of shunting quite high currents. The second point appears not because the system is routinely operated with a large excess power capacity, which would be symptomatic of a badly designed link. The large current capability is necessary because the transferred power is very sensitive to movement of the transmitter relative to the receiver, and it can become an order of magnitude larger than the nominal value with only small position shifts.

The maximum gate-source voltage on the shunt transistor in figure 4.6 is approximately one threshold voltage over the bandgap voltage V_{REF}. This is not quite true since the differential pair is biased in weak inversion, but on the other hand, the saturation voltage of transistors in weak inversion is only a few times the thermal voltage U_T [116]. With a reference voltage of 1.25 V the linear output range of the amplifier extends up to around 1.70 V. The upper limit of the output range is set by the minimum $|V_{DS}|$ across M2. To get the maximal gain-bandwidth out of the amplifier the gate length of the shunt transistor is set to the minimum. The width is then selected to pass the specified maximum shunt current at a gate drive of 1.70 V. In 0.5 µm and smaller technologies, short-channel effects have to be included when calculating the drain current at such a high effective gate voltage. The principal effect in this case is velocity saturation. In the presence of velocity saturation, the drain current becomes [63]

$$I_d = \frac{K_N' W}{2L}(V_{gs} - V_t)\left[(V_{gs} - V_t)||LE_{sat}\right] \tag{4.12}$$

where E_{sat} is the saturation field strength. This is the electric field strength at which the carrier velocity is half the value extrapolated the from weak-field mobility. When the ratio of LE_{sat} to $(V_{gs} - V_t)$ is small, the traditional second-order characteristic can be used. The saturation field strength of the technology is not specified, but a reasonable estimate for a 0.5 µm technology is 4×10^6 V/m [63]. In that case, the drain current at 1.1 V gate overdrive is 35% lower than it would be without velocity saturation. In the following, this effect will be neglected, both because the real saturation velocity is not known, and because the error is similar to those that can be expected from normal process variations.

Using a simple quadratic expression for the drain current, the minimum width of the shunt transistor is

$$W = \frac{2LI_{d1,max}}{K_N'(V_{gs1,max} - V_t)^2} \tag{4.13}$$

With a maximum current of 50 mA and a gate voltage overdrive of 1.1 V, the minimum width of the shunt transistor is 320 µm. Adding the effect of velocity saturation with the E_{sat} given above, the minimum width becomes 490 µm. The shunt transistor in this incarnation of the regulator was given a width of 600 µm after application of a fudge factor, so that it would handle the rated current of 50 mA despite possible processing surprises.

The maximum current density in transistor channels and diffusion are usually not specified in the wafer manufacturer's electrical rules. What is available are the electromigration limits in conductor layers. Electromigration is not the problem when the shunt transistor is dimensioned, but rather power dissipation. The transistor is laid out as a compact finger structure to reduce the total capacitances, and this makes for a relatively small area of $30\times50\mu m^2$. At a shunt current of 50 mA the transistor dissipates around 165 mW. While this is not enough to raise the surface temperature by more than 10–20°C, the temperature gradient in the neighborhood of the device can affect other analog circuits. The temperature rise and the magnitude of the gradients can be checked using a heat transfer simulation program to calculate how fast heat is transferred away through the silicon from a surface source with the same area as the transistor.

As we mentioned above, the gain-bandwidth of the amplifier is set both by intrinsic properties of the amplifier and the load capacitance. The amplifier is a simple differential pair with both the positive and negative branch mirrored to the output. Because of this, the transconductance of the stage is equal to the transconductance of the input pair transistors g_{m2}. The input pair is laid out using wide transistors to achieve weak inversion operation, for maximum gain per unit bias current. The offset error of the differential pair is highest in weak inversion [116], but that is not important in this case, since even an offset voltage of 10 mV only leads to a shunt threshold error of 25 mV. Setting the capacitance looking into the gate of M1 to C_g, the GBW of the opamp is

$$\omega_1 = \frac{I_T}{2nU_T C_g} \tag{4.14}$$

where n is the usual subthreshold slope factor for the input transistors. The calculation of the load capacitance is complicated by the fact that the shunt transistor is in weak inversion over a large range of operating currents. Because of its large W/L ratio, M1 is in weak or moderate inversion up to a drain current of approximately 1 mA. This covers a large part of the expected normal operating range. In weak and moderate inversion, the small-signal gate capacitance can be up to several times smaller than the accumulation and strong inversion value (see for example [109]). The exact behavior of the capacitance as a function of gate-bulk voltage is quite complex, and depends on some unpublished process parameters. In the interest of conservative design, it is however possible to simply use the largest value of the capacitance, the strong inversion value $C_{ox} = \varepsilon_s/t_{ox}$. A lower load capacitance increases the GBW of the amplifier, and (4.9) shows that an increased GBW improves the stability of the regulator. Using the conservative expression for the gate-source capacitance we have

$$C_g = C_{gs} + (1 + g_{m1}Z_d)C_{gd} \tag{4.15}$$

The small-signal impedance level Z_d seen by the drain terminal is equal to

$$Z_d = \left(\frac{1}{\beta g_{m1} A(s)}\right) || \left(\frac{1+sRC}{sC}\right) \tag{4.16}$$

(the feedback loop must be closed for this calculation). As long as the frequency $1/RC$ is much smaller than the GBW, this impedance is much smaller than $1/\beta g_{m1}$ for all frequencies. The Miller feedback term in (4.15) can then be neglected, and the load capacitance is

$$C_g = \frac{2}{3}WLC_{ox} + 2WC_{ov}. \tag{4.17}$$

We had already obtained an expression for the transistor width as a function of the maximum shunt current in (4.13). Using this along with the GBW (4.14), the GBW can be written as a function which is simply the ratio between the tail current and the maximum shunt current, multiplied by a factor which depends only on processing parameters and the gate overdrive $V_{emax} = (V_{gs,max} - V_t)$ of M1 (which is set by the reference voltage and the topology of the regulator):

$$\omega_1 = \frac{1}{2nU_T}\left[\frac{4}{3}C_{ox}\frac{L^2}{K'_N V_{emax}^2} + 4C_{ov}\frac{L}{K'_N V_{emax}^2}\right]^{-1}\frac{I_T}{I_{d1,max}}. \tag{4.18}$$

Note that this ratio also sets an upper limit to the range of shunt currents that the regulator can operate with. For a given ω_1, the minimum current through the regulator (neglecting the voltage reference and biasing) is I_T and the maximum is $I_{d1,max}$.

The damping coefficient equation (4.9) can be used to fine-tune the design variables to obtain stable operation over the entire operating current range. The equation

$$\zeta = \frac{1}{2A_{dc}}\sqrt{\frac{C\omega_1}{g_{m1}\beta}}(1 + Rg_{m1}\beta A_{dc}) \tag{4.19}$$

may be rewritten as

$$\zeta = \frac{\Gamma}{\sqrt{g_{m1}}} + \Lambda\sqrt{g_{m1}} \tag{4.20}$$

with

$$\Gamma = \frac{1}{2A_{dc}}\sqrt{\frac{C\omega_1}{\beta}} \tag{4.21}$$

$$\Lambda = \frac{R}{2}\sqrt{C\omega_1\beta} \tag{4.22}$$

It is easiest to picture the two terms in (4.20) as two straight lines on a log-log graph, whose intersection lies at the point where $g_{m1} = \Gamma/\Lambda = 1/(\beta A_{dc}R)$. This is also the point where their sum ζ has a minimum. We assume that the variables A_{dc}, β and C have already been set by other constraints. There remains to be found a combination of R and ω_1 which gives a satisfactory value of DC output resistance and a damping coefficient ζ greater than a given minimum. The gain-bandwidth ω_1 appears in both terms, so the damping coefficient scales as the square root of ω_1. The resistance R only appears in the second term, so it can be used to increase the damping at high shunt currents. The effect of R upon damping at low shunt currents is however negligible. Figure 4.7 shows a graph of the damping coefficient as a function of g_{m1} for a fixed resistance R, parameterized with the gain-bandwidth ω_1. Figure 4.8 is a complementary graph, with the gain-bandwidth fixed and variable R. These figures can be used to add weight to the claim that a compensation resistor is necessary. Without it, the $\zeta(g_{m1})$ function would not have a minimum, and would decrease toward zero for large g_{m1}. It would then be necessary to increase ω_1 to impractical levels to obtain good damping across the whole operating region.

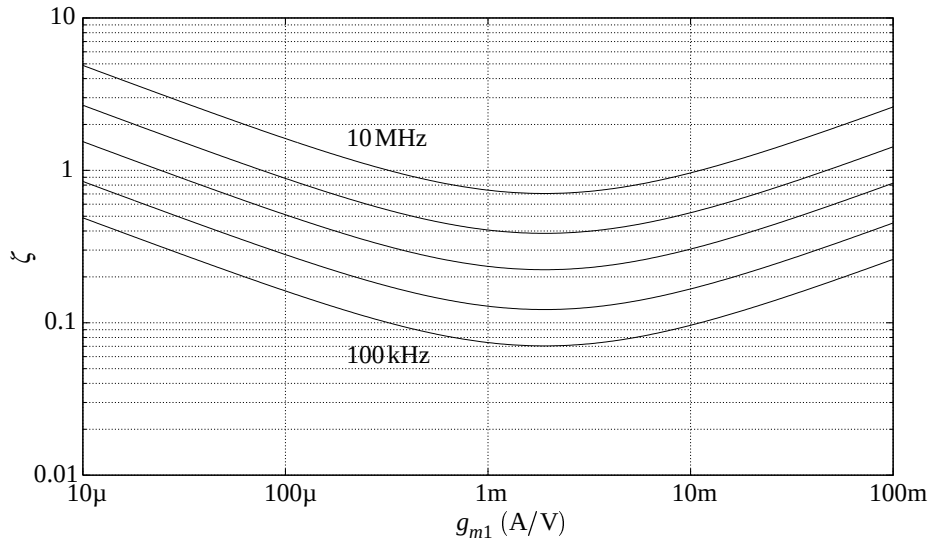


Figure 4.7: *The regulator damping coefficient as a function of bias point for $R = 5\,\Omega$, and variable amplifier gain-bandwidth $\omega_1 = 0.1, 0.3, 1, 3, 10\,MHz$. $A_{dc} = 50\,dB$, $\beta = 1/3$, $C = 500\,nF$.*

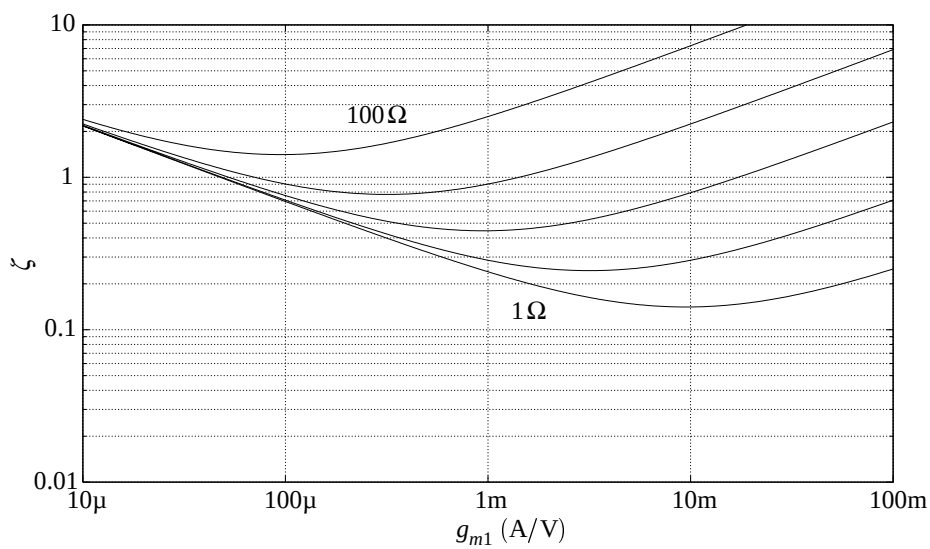


Figure 4.8: *The damping coefficient for a gain-bandwidth of 2MHz and variable compensating resistance $R = 1, 3, 10, 30, 100\,\Omega$. $A_{dc} = 50\,dB$, $\beta = 1/3$, $C = 500\,nF$.*

4.1.3 Design considerations

The load capacitance presented by the gate of the shunt transistor turns out to be around 0.8 pF in the strong inversion limit. This fact can now be used in conjunction with equations

(4.14) to deduce that the tail current in the differential amplifier needs to be 1.0 µA. The DC gain of the amplifier is as usual equal to the stage transconductance times the small-signal impedance level at the output node. Because the current mirror transistors M4 and M5 were laid out with long gates, their r_{ds} output resistance is several times greater than the output resistance of the input transistors. Ignoring the contribution of M4 and M5 to the output resistance, the gain of the amplifier is seen to be equal to the intrinsic gain of the input transistors

$$A_{dc} = \frac{g_{m2}}{g_{ds2}} \tag{4.23}$$

Calculating the output conductance of transistors using modern technologies and fabrication data can be problematic. On the one hand, sources which attempt to treat the transistor behavior analytically [109, 34] rely on process data such as doping levels and diffusion depth, which are rarely published by process providers. On the other hand, there are the semi-empirical models which are currently provided with most processes, like BSIM3v3 [36, 111]. These may provide good simulation accuracy for transistors operating in the modeled regime, but do not extrapolate well to other operating conditions. The BSIM parameters provided with this process were clearly extracted from strong-inversion transistor data, as they gave quite inaccurate results for weak inversion. These inaccuracies can be verified for example by simulating circuits whose outputs are only functions of verifiable physical quantities, like the constant-g_m bias cell [63].

The EKV MOS transistor model [34, 36], which attempts to model the MOS transistor in all operating regions, can provide more accurate simulation results in cases like this. Few foundries do however provide EKV parameters for their processes at present. The channel length modulation in the EKV model depends on a fitted parameter, so without this model-specific process parameter the model cannot be applied to calculate the output resistance of a transistor current source.

Falling back on the classical channel length modulation model, we have

$$I_d = I_d|_{V_{dsat}}(1 + \lambda V_{ds}) \tag{4.24}$$

with the λ factor dependent on the carrier concentration, among other things [55]. Experimental results show that this factor is largest in weak inversion, and it tends to decrease somewhat in strong inversion [116]. Using this channel modulation model, the conductance g_{ds} is proportional to the drain current and the amplifier gain (4.23) is

$$A_{dc} = \frac{1}{nU_T\lambda} \tag{4.25}$$

The nondominant pole originates in the node capacitance and impedance of the input to the current mirror M4-M5. As is well known, a pole of this type only acts on half the signal, and therefore generates a pole at the frequency and a zero at twice the frequency [59]. In a unity-gain feedback configuration, it would be necessary to place the nondominant pole at two or three time the GBW, to provide an acceptable phase margin. In our case, the gain in the feedback branch is significantly smaller than one. As was pointed out in the discussion about the Miller capacitance of M1, the voltage gain of M1 is smaller than unity. The feedback branch multiplies this by an additional factor $\beta \approx 0.3$. The nondominant pole turns out to have a frequency of $4\omega_1$, but even significantly lower values no not degrade the overall stability.

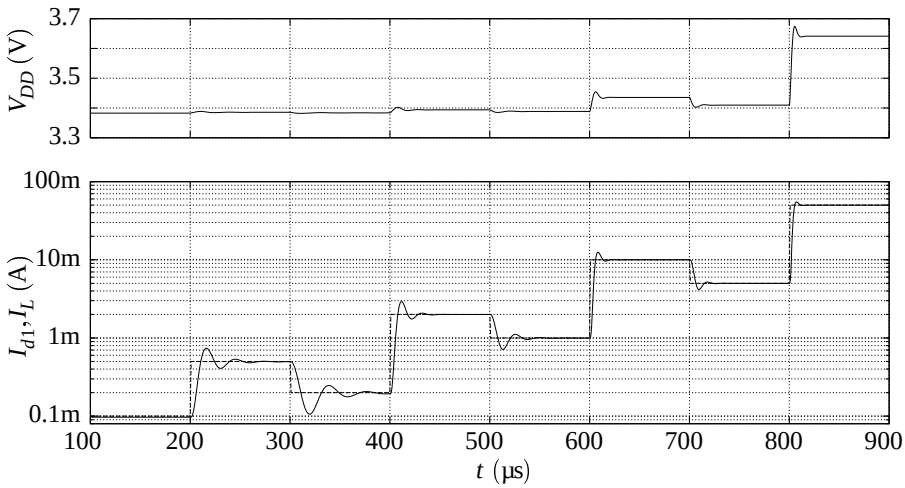


Figure 4.9: *Transient simulation results for the shunt regulator with a logarithmic rebated staircase input current.*

The resistance R was set to 5 Ω and the GBW to 2.5 MHz. This gives a damping coefficient which is 0.2 or larger for all operating currents. A larger minimum damping coefficient could have been obtained by spending more current in the operational amplifier for a larger GBW, but it was not considered necessary.

The current slew rate of the shunt regulator could in some cases be insufficient, mainly if it were not capable of tracking power transfer variations due to fast changes in transmitter-receiver geometry. The current slew rate is related to the slew rate $\mathrm{SR} = I_T/C_g$ of the amplifier output by the relation

$$\mathrm{SR}_I = \frac{g_{m1} I_T}{C_g} \tag{4.26}$$

Since the shunt transistor is in strong inversion at the higher power levels, this can be rewritten as

$$\mathrm{SR}_I = \frac{I_T}{C_g}\sqrt{\frac{2 I_{d1} W}{L}} \tag{4.27}$$

Any variation in excess current that is not absorbed immediately by the shunt is diverted into the supply capacitor C. As a worst case example, we can assume that the power transfer change is an instantaneous step change. An interesting measure of the regulator's performance is then by how much the capacitor voltage changes while the regulator is catching up. In practice, the current slewing is fast enough to track any rate of load change that is likely to occur. A worst case estimate of the time needed to switch from zero to the rated current is $t_r = (C_g/I_T)V_{gs1,max}$. With the values used in the design of the regulator, this is $t_r = 1.4\,\mu\mathrm{s}$. The response is then faster than any conceivable input power changes under normal operating conditions. A simulation of the transient response of the regulator is shown in figure 4.9. The applied load current is a piecewise linear function which shows the step response for a number of operating points in the range 100 µA–50 mA.

Most of the die space of $300\times120\,\mu m^2$ used by the shunt regulator is taken up by the voltage reference. The $5\,\Omega$ resistor R is laid out in normal low-resistance salicided poly because of its low value. The area of the resistor is purposely made large, for two reasons. A large width for a given length/width ratio reduces the current density in the material, and permits a better contact layout at the endpoints. A large area also reduces the power density at a given power level. At the rated shunt current, the resistor must dissipate around 13mW. While this is not very much, it can establish a heat gradient in the substrate that can disturb nearby precision circuits.

The most concentrated heat source is the shunt transistor itself. At the nominal current of 50 mA, the power dissipation is close to 170 mW. As the test results in the following section show, the actual implementation can handle up to 250 mW. This is enough to change the substrate temperature significantly in the neighborhood of the regulator. The regulator (like the other power supply components) must therefore be placed at the edge of the chip, as far away as possible from other analog circuits.

4.1.4 Figure of merit

There are clearly many tradeoffs involved in the design of this regulator and to obtain an overview of the possible choices it can be useful to define a figure of merit for the design. Such a figure also makes it easier to compare the performance of this specific regulator to other designs. An obvious choice for the figure of merit is the ratio between the minimum and maximum currents through the shunt. The range of shunt currents is more meaningful in this context than the maximum current since the latter can be increased arbitrarily by simple scaling. Some important performance figures like the damping coefficient and output resistance can be traded off for a larger current range, so the FOM should be specified for a set of fixed values of these.

The minimum shunt current consists of the tail current of the differential pair, I_T, and the bias current I_B for the tail current source and through the voltage reference. We chose to design the shunt as a self-contained unit but a voltage reference and bias voltage sources are usually available on a system chip. The bias current part I_B was therefore left out of the definition of the FOM, because this regulator is designed for integration in an implanted system, and because the resulting expression for the FOM becomes much simpler. The maximum shunt current is essentially the drain current of the shunt transistor M1 at the maximum gate bias, since it is presumably orders of magnitude larger than I_T or I_B. The figure of merit is then

$$F = \frac{I_{d1,max} + I_T}{I_T} \approx \frac{I_{d1,max}}{I_T} \tag{4.28}$$

As constraints for the FOM, we choose to fix the minimum damping coefficient at a value of 0.5, and set the output resistance to such a value that the supply voltage varies by 10% over the shunt current range. The minimum of the damping coefficient (4.20) occurs at $g_{m1} = 1/(\beta A_{dc} R)$ and has the value

$$\zeta_{min} = \sqrt{\frac{C\omega_1 R}{A_{dc}}} \tag{4.29}$$

The gain-bandwidth of the amplifier with the shunt transistor gate load in (4.14) is proportional to $I_T/I_{d1,max}$. The proportionality constant depends on process parameters, the temperature and the reference voltage, and we assign it the name Π. If the DC gain of the amplifier is $A_{dc} = 1/(nU_T\lambda)$, we then have

$$\zeta_{min} = \sqrt{\frac{CR\Pi I_T}{A_{dc} I_{d1,max}}} \tag{4.30}$$

Rearranging and inserting the expressions for A_{dc} and Π, we get

$$F = \frac{I_{d1,max}}{I_T} = \frac{CR\lambda}{2\zeta_{min}^2} \frac{3K_N'(V_{gs1,max} - V_t)^2}{4C_{ox}L^2 + 12C_{ov}L} \tag{4.31}$$

This expression reveals which parameters influence the achievable current range. There are the fixed technology parameters C_{ox}, C_{ov}, K_N' V_t, λ and L (which is equal to L_{min}), and the system design parameters C, V_{REF}, ζ_{min} and R.

The channel-length modulation parameter λ can be increased, thus increasing the gain of the amplifier, by using non-minimum length transistors in the differential pair. This can only be exploited to a limited extent since the width of the transistors must be scaled accordingly to maintain weak inversion, and at some point the nondominant pole of the amplifier will move too close to the dominant pole. Some extra performance can still be won by using more than minimum gate length in the differential pair.

The suggested design procedure for obtaining the maximum current range given $I_{d1,max}$, C, ζ_{min} and the allowed supply voltage variation is as follows. The compensation resistor is set to $R = \Delta V_{DD}/I_{d1,max}$ or a little less to allow for finite amplifier gain. The tail current can then be determined directly from (4.31). As an example we cant take the implemented shunt regulator, with a supply voltage limit of 3 V, $C = 500$ nF, $\zeta_{min} = 0.2$, a maximum shunt current of 100 mA and an allowed supply voltage variation of 10%. The value of the compensation resistor is then $R = 3\Omega$, and plugging in technology data for the 0.5 µm process results in a figure of merit, or shunt current range, which is $F = 3.0 \times 10^5$. The tail current is then $I_T = 0.3$ µA.

4.1.5 Shunt test results

The regulator was implemented as a separate two-terminal device, with its own voltage reference, partly to simplify testing. It could therefore be tested separately, without powering the rest of the system that it was a part of. The "load current" in the diagrams above could therefore be controlled explicitly during testing. The transfer function of the regulator was tested simply by adding an external series resistor and a variable voltage source, and inferring the shunt current from the voltage drop over the resistor. The AC measurements were performed in the same way, simply adding an AC component to the applied DC voltage.

Figure 4.10 shows the results of the DC characteristic for one sample of the regulator. The graphs have several interesting features. The first plateau reached by the shunt current is the regulator's own current consumption of 3.0 µA which does not lie far from the design value of 2.8 µA. The knee voltage is at 3.45 V instead of the simulated value of 3.38 V. The difference has three likely causes, which are in order of increasing importance the opamp

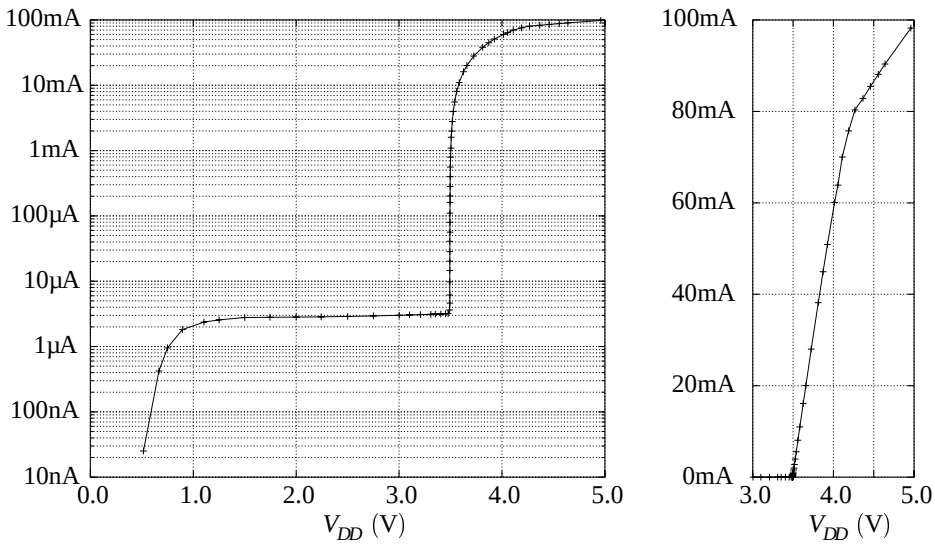


Figure 4.10: *Experimental data for the current through the shunt regulator as a function of supply voltage. The data is shown both on a log-log scale and a linear scale.*

offset, reference error and inaccuracy in the active divider. The high knee voltage means that the supply voltage exceeds the absolute maximum voltage for the technology, which is 3.3 V+10%, when the shunt current is close to its upper limit. For a production system, the knee voltage would have to be set more conservatively, to avoid the reduction in system lifetime that is caused by operation above the rated supply voltage.

The output resistance of the regulator is the slope of the measured transfer characteristic. It is shown in figure 4.11 as a function of the DC shunt current. The large uncertainty in the measured resistance at small current levels is due to the difficulty in measuring absolute voltages with the required accuracy. The change in voltage across the regulator for the two first points in figure 4.11 is only 50ppm, so voltage changes at two consecutive measurement points can easily disappear in power supply and voltmeter drift.

The resistance is around 8 Ω over most of the operating range. At low bias currents the resistance is somewhat higher because of the reduced loop gain. The operating range actually extends up to 80 mA or so, instead of the nominal 50 mA. The generous dimensioning of the shunt transistor M1 is the main reason for the difference. The resistance increases sharply again at the upper extreme of the operating range, as the opamp transistor M2 leaves saturation. This causes the loop gain to drop, with a resulting increase in output resistance.

There were 10 samples of the chip available for testing. They had knee voltages (defined here as the supply voltage causing a shunt current of 20 µA) ranging from 3.42 to 3.49 V, with an average value of 3.45 V. Because of the small sample size, more elaborate statistics for the measured distribution are meaningless. According to the fabricator's data, all chips came off the same wafer, so we are not seeing the full range of process variations.

The AC output impedance of the regulator was also measured, and the results are shown in figure 4.12. Because of the small amplitude of the measured voltages and low oscilloscope sensitivity, it proved impossible to measure the impedance at lower DC bias levels. In order

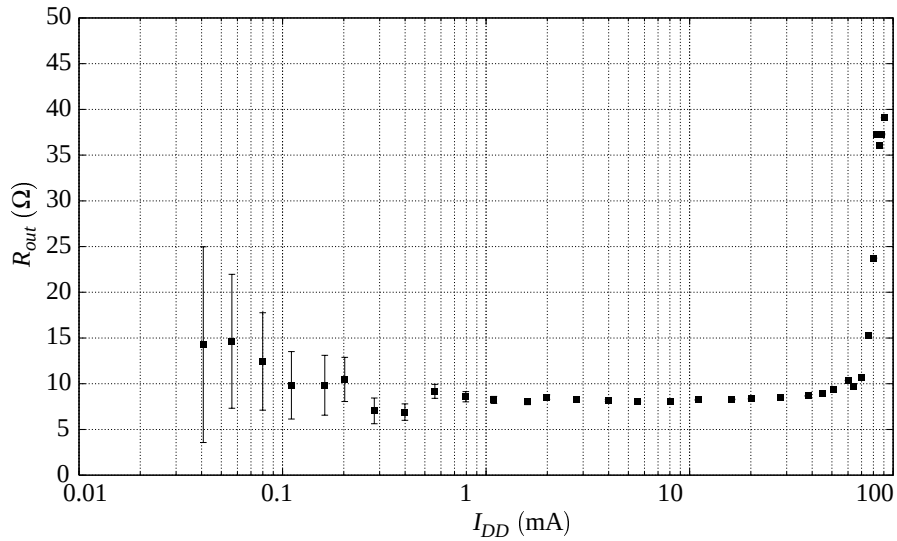


Figure 4.11: *The measured DC small-signal output resistance of the regulator as a function of bias point.*

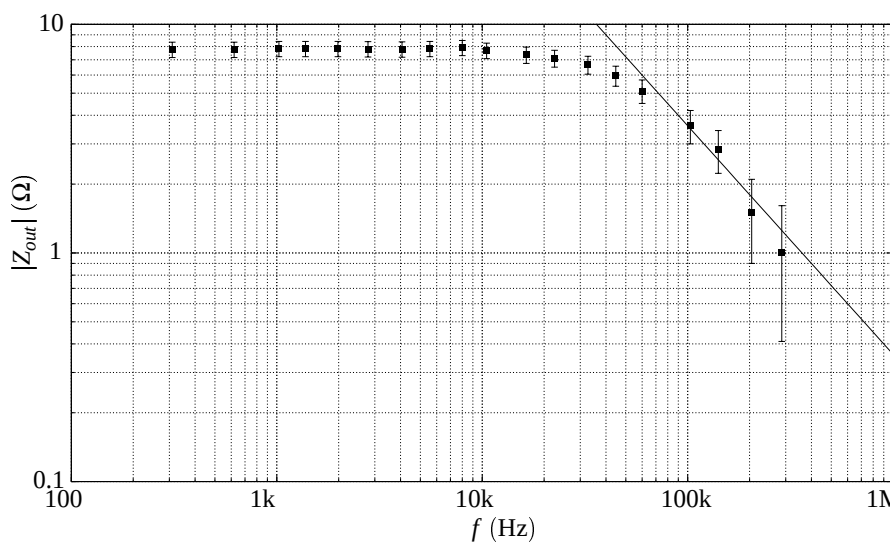


Figure 4.12: *Measured AC impedance at a DC bias of 20 mA. The external capacitance was* $C = 540\,nF$*. The figure also shows a -20 dB/decade asymptote.*

to satisfy the small-signal requirement, the AC signal was kept at 10% of the DC bias. For a DC bias of 20 mA, the maximum allowed signal was then $2\,\mathrm{mA_{p\text{-}p}}$, which gives a voltage amplitude of $16\,\mathrm{mV_{p\text{-}p}}$ over an $8\,\Omega$ resistance at DC, with the amplitude decreasing at higher

frequencies.

The tail current in the differential pair was set to 0.5 µA in this implementation. The performance of the shunt regulator is quite impressive considering this, with the measured shunt current changing by a factor of 10^4 over a 47 mV supply voltage range, and by a factor 10^5 over a 10% increase in the supply. Even if the bias and reference currents are included in the fixed part of the shunt current, the performance is still respectable.

4.2 Rectifiers

The rectifier or power converter is the interface between the external RF signal and the internal power supply of the system. Since the quality of the power supply is influenced to a great degree by the signals passing through the rectifier it is necessary to examine in detail how system specifications affect the dimensions and performance of the rectifying elements. The main performance parameters related to the converter design are power dissipation in the rectifiers, internal supply noise, and loading of the external signal. All these parameters depend both on the topology of the converter (half bridge, full bridge etc.) and on the nature of the rectifiers.

The range of available rectifier elements is technology-dependent, assuming that on-chip rectifiers are used. External rectifiers are of limited interest when on-chip solutions are available, since they increase the bulk, cost and interconnect complexity, and they may reduce the reliability of the system. The design of on-chip rectifiers with acceptable performance can be quite challenging, and in the special case of a pure CMOS technology, the lack of isolated p-n junctions dictates the use of other means.

There are two main fields of application for combined power and data transmission to integrated circuits. These are of course the field of biomedical telemetry, and the various applications of contactless smart cards [44, 17, 70]. The power conversion conditions prevalent in the latter case are somewhat different from the conditions for biomedical applications. The available transmitter power is usually far greater since the card readers are fixed stations which do not operate off portable power sources. The nature of the on-chip signal processing is also different. In general, contactless smart cards contain only digital signal processing, which relaxes the requirements of power supply quality. Another important difference is the efficiency of the power conversion circuits in the smart card circuit or the implanted biomedical system. Due to the difference in available transmitter power, the biomedical system designer is forced to use more efficient power conversion topologies, and more efficient circuit elements to implement the chosen topology. The efficiency of the circuit elements (which are typically rectifiers in this context) refers to the power dissipation in the elements themselves, and for elements which suffer from parasitic circuit devices, on how much current is diverted from the desired path.

One of the earliest implementations of a contactless smart card power conversion circuit which was fully integrated in a plain CMOS process is shown in [44]. It is based on the use of NMOS transistors (in a common p-well), connected as a full bridge rectifier. The circuit was operated with a carrier of 1 MHz and actually had the receiver antenna (the secondary inductor/coil) integrated in metal on the chip itself. This is possible with the field intensity and small separation available in a smart card apparatus. The effect of parasitic elements in the rectifier bridge was however not mentioned except in passing, although they are quite

important, especially in a technology with threshold voltages on the order of 1 V or higher.

More recent published results regarding integrated power conversion in smart card circuits show BiCMOS full bridge rectifiers operating at 13.56 MHz (one of the ISM bands) [70], and a full bridge CMOS rectifier operating at 5 MHz using both NMOS and PMOS elements [17]. The performance deterioration due to the parasitics seems to be largely ignored in the smart card literature, a pardonable omission since the operating regime of the external equipment can be adjusted to compensate for this without any significant penalty.

Applications of power conversion in the biomedical telemetry field differ from corresponding applications for smart cards in that more attention is paid to efficiency and the quality of the on-chip power supply. The rectifier elements are almost without exception implemented using p-n junctions in BiCMOS processes [50, 122, 4, 68] or with external (off-chip) discrete components [98]. The converter topologies are in most cases full bridge rectifiers, with a few exceptions. In [50] for example, a voltage-doubling (Villard) converter is used because the sensor circuitry must operate from the unusually high supply voltage of 12 V. [4] uses a bridge rectifier with 2 diodes in series in each branch because of the limited breakdown voltage of the p-n junctions.

In the rest of this section, the different converter topologies will be examined with respect to efficiency, supply quality and signal quality, and we will go on to demonstrate a plain CMOS bridge converter which does not suffer from the common parasitic degradation problems.

The discussion which follows is of a general nature, and can be applied to most operating regimes. Where numerical examples are shown however, a carrier frequency of 7 MHz was used (chosen for numerical convenience, but applicable to the 6.78 MHz ISM band) and an internal supply voltage of 3 V.

4.2.1 Topologies

There are several different circuit topologies that can be used to convert the RF input power to a usable DC supply voltage. The choice of conversion circuit depends on the application, but in the end, it reduces to a compromise between power dissipation, area, input loading and supply noise spectrum. Five representative circuits are shown in figure 4.13, which cover the range of practical possibilities. The rectifier elements are represented by diode symbols, but they can be any kind of device with a preferred current direction. The exact rectifier type is not important for this circuit-level overview, but the possible rectifier implementations are shown in a later section.

The circuits have a common trait in that they only conduct current from the input LC tank during a small fraction of the tank period. This simplifies the estimation of the power dissipation, since we can assume with no great loss of precision that the entire supply current is transmitted while the rectifier forward drop V_F is maximum. This approximation is justified if the forward current in the rectifiers increases "rapidly" as a function of the forward voltage, that is exponentially (as for a p-n junction) or as a power function. The power dissipation in the rectifiers can then be estimated quite simply. If the DC supply current through the aggregate load Z_L is I_{DC}, the power dissipated in the rectifiers is $P_D = I_{DC}V_F$ in the half-wave circuits, and $2I_{DC}V_F$ for the full bridge. In that respect, the full bridge rectifier is an inferior solution, especially if the forward voltage drops V_F are large compared to the supply voltage. In general, the conversion efficiency, neglecting resistive losses in the LC

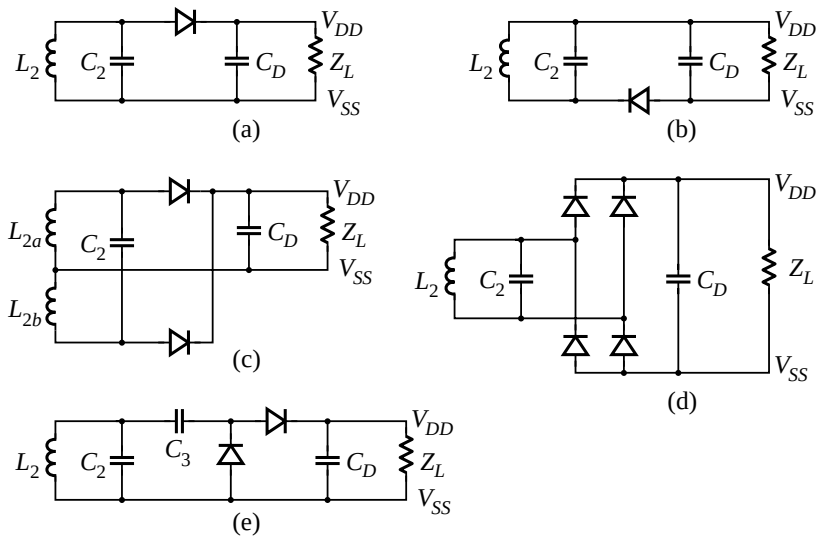


Figure 4.13: *Different power conversion arrangements. (a) and (b) show basic half-wave rectifiers, with different rectifier connections. (c) shows a full-wave rectifier, (d) a full-wave bridge rectifier, and (e) a half-wave voltage doubling rectifier.*

tank and the rectifiers, is

$$\eta = \frac{I_{DC}V_{DD}}{P_D + I_{DC}V_{DD}} = \frac{V_{DD}}{mV_F + V_{DD}} \tag{4.32}$$

where the number m is 1 or 2 depending on the configuration.

The half-wave rectifier in figure 4.13 is shown with two possible rectifier connections, one with a rectifier terminal connected to V_{DD} and one with a terminal connected to V_{SS}. The full-wave rectifier in (c) can also be implemented with both types of connection. This can be an advantage when it comes to the implementation of the rectifiers since one type may be of much higher quality than the other. This is especially true for spartan technologies like digital CMOS processes.

The spectrum of the current through the power conversion into the supply voltage node depends mostly on whether the rectifier is full-wave or half-wave. In the former case, the current pulses have roughly half the amplitude and have frequency components at twice the frequency of the latter case. This is a clear advantage of full-wave rectification, since the impedance of the supply voltage node is mostly capacitive at high frequencies. All other things being equal, the lower current amplitude and higher frequency are then both characteristics which tend to lower the amplitude of the induced supply voltage transients.

The respective areas of the rectifier implementations are not an important consideration if high-quality p-n junction diodes are available. The current density in that type of rectifier is high enough, that any conceivable supply current (up to 100 mA for example) can be accommodated in a relatively small area. More limited technologies like plain CMOS do however have limitations which make the rectifiers consume a lot more area. A short area comparison of the topologies is therefore in order. Let us assume that the necessary area

for a rectifier scales linearly with the peak current. The rectifier areas for the half-wave and full-wave topologies in figures 4.13a and 4.13c are then equal because of the difference in peak current. The full-wave bridge in 4.13d on the other hand, needs twice the area.

Except in cases where a separate data channel is used, the antenna signal serves to transfer modulated data to the system. The nature of the load seen by the antenna can therefore have a large influence on the feasibility of different data transmission schemes. In general, it is the high nonlinearity of the load that is the problem. The power conversion circuits shown above act as soft limiters on the antenna signal, leading to various nonlinear effects, with intermodulation being the most undesirable. The harmonic distortion can be ignored in basic envelope detector demodulator designs. The half-wave rectifier load is asymmetric, so even-order distortion products appear in the signal. These can degrade the performance of direct-conversion receivers [1, 15]. The full-wave and bridge converters present a symmetric load to the antenna, so the odd-order distortion should dominate (assuming that the signal in the full-wave case is the differential signal across both inductors). The full-wave converter does however cause some differential-to-common mode conversion, which may be a problem depending on the receiver.

To recapitulate, we can list the arguments for and against using each type of power converter. One virtue of the half-wave rectifier is its simplicity, but while this property may facilitate the analysis of the circuit, it does not have any practical value for the implementation. More complex implementations are preferable if they provide better performance at no increased cost in area, power or interconnect complexity. Another point in favor of this converter which has a more practical value is that the power dissipation is minimal. The downside of the half-wave converter is the increased current pulse amplitude, the lower current pulse frequency, the asymmetric loading of the tank circuit and the lack of a differential signal tap.

The full-wave rectifier uses the same total area as the half-wave version, but has better transient current characteristics and the same power dissipation. Both the half- and full-wave converters can be implemented using rectifiers which have one terminal fixed at V_{DD} or V_{SS}, a distinct advantage with some technologies. The full-wave converter does however require a more complex antenna, with a central tap on the inductor. It does have a differential signal output, taken across the entire antenna. In a rather unusual reversal of the normal situation, the large amplitude of the signal (over two times the supply voltage) makes it necessary to insert an attenuator before the receiver. This is actually the case with all the power converters if a shared power/data link is used.

The full-wave bridge converter does not require any additional external components or connection compared to the simple half-wave converter. The area requirement is however two times as high, and two rectifier polarities must be used (to V_{DD} and V_{SS}). The power dissipation is also two times as high. If the area and power costs are acceptable, it offers a supply current spectrum which is similar to that of the full-wave converter, without the external antenna and interconnect complexity. It also provides an input signal with a minimum of even-order distortion and a differential signal with a reduced amplitude and a minimum common-mode component.

The voltage doubling converter needs a large external capacitor in series with the antenna. The spectrum of the current that is passed through the converter is similar to the case of the half-wave converter, since there is only one current pulse per carrier period. The power dissipation of the doubling converter is however closer to the dissipation of the full-

wave converters since the full supply current is passed through a rectifier twice per period. For the same reason, the area requirement for the rectifiers is two times that of the half- and full-wave converters. For these reasons, the voltage doubling converter should only be used when the supply voltage requirement can not be met by the other types.

The best converter for a given system depends on which of the conflicting performance specifications is the most important one. In general, lower supply transient currents are desirable, but this may have to be traded off against the low power dissipation of the half-wave converter coupled with its simple antenna construction. Receiver requirements may on the other hand dictate a full-wave conversion, bridge or not. In the following, a full-wave bridge converter is used as a reference case, since it is the most challenging to implement and provides the best signal characteristics. The other converter types can then be regarded as a subset of the full bridge, and the results obtained for the bridge can be applied to them.

Rectifier power dissipation

The argument for the power dissipation approximation used above was that the dissipation for some rectifier types is close to the average current multiplied by the peak voltage drop V_{PK}. This applies to devices for which the current increases rapidly as a function of the voltage drop, specifically as an exponential or power function. To provide some support to this claim, a brief treatment of the exponential and power law behavior is given for a representative pulse shape.

The relation between the forward drop $V_{D1}(t)$ and the current $I(t)$ through a p-n junction can be expressed as

$$V_{D1}(t) = U_T \ln \frac{I(t)}{I_S} \tag{4.33}$$

For a rectifier with a power law characteristic (like a MOS rectifier in strong inversion, with $n = 2$) the corresponding expression is

$$V_{D2}(t) = U_2 \left(\frac{I(t)}{I_2} \right)^{1/n} \tag{4.34}$$

where U_2 and I_2 are the coordinates of any point on the curve. The error in the constant voltage drop approximation can be estimated by comparing the integral of the instantaneous power $P(t) = V(t)I(t)$ over one current pulse to the approximate energy dissipation $W_a = V_{PK} \int I(t)\,dt$.

A triangular current pulse with peak current I_{PK} and width T_p is used in this example since it is one of few shapes which yield analytic results. The power dissipation integral for the p-n junction case is then

$$W_1 = 2 \int_0^{T_p/2} I(t) v_{D1}(t)\,dt \;=\; \frac{4U_T i_{PK}}{T_p} \int_0^{T_p/2} t \ln \left(\frac{2i_{PK}}{I_S T_p} t \right) dt \tag{4.35}$$

The solution to the integral is found by substitution and partial integration with the result

$$W_1 = \left[U_T \ln \left(\frac{i_{PK}}{I_S} \right) - \frac{U_T}{2} \right] \frac{i_{PK} T_p}{2} \tag{4.36}$$

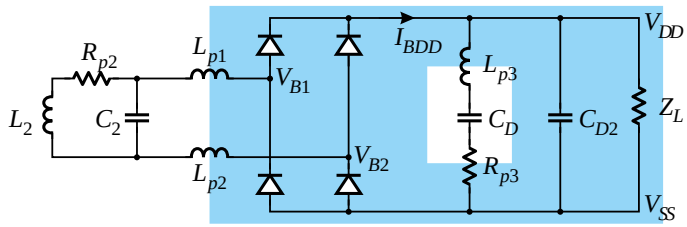


Figure 4.14: *A simple bridge rectifier along with the most important parasitic elements. The shaded area marks the on-chip elements.*

The first term inside the brackets is the peak voltage drop and the second term is a small correction factor of half a thermal voltage. The factor $i_{PK}T_p/2$ is the total charge through the device, and multiplied by the contents of the brackets it gives the total energy dissipated in the rectifier during one current pulse. The sum inside the brackets is therefore the effective average voltage drop over the rectifier, and the error in our approximation is simply $U_T/2$. This correction factor is only a few percent of v_{PK} under normal operating conditions (peak forward drop greater than 500 mV) so our approximation is fully justified.

For a device with a power law forward characteristic the result of the integral is

$$W_2 = v_{PK}\frac{i_{PK}T_p}{2}\frac{2n}{2n+1} \tag{4.37}$$

This can be applied to a MOS diode rectifier with $n = 2$ regardless of the offset in the characteristic due to the threshold voltage. The consequence of the threshold voltage is that the relative error due to the approximation is smaller for a MOS than in (4.37). As expected, the error in the power dissipation approximation becomes smaller as n increases. The estimate is too high by the factor $(2n+1)/2n$ which is 25%, 16% and 11% too high for $n = 2$, 3 and 4 respectively.

These results were derived for triangular current pulses but other shapes give similar results. The approximation is exact for rectangular pulses, while other shapes lead to varying degrees of error. There is however no point in pursuing this further by attempting to analyze other pulse shapes exactly since they will probably not come closer to the true pulse shape than the triangular approximation. The point made here is that the power dissipation can be estimated quite well in this simple manner, certainly well enough to compare different power converter topologies based on identical rectifier types.

4.2.2 Dynamic full bridge behavior

The rectifier is the interface between the external RF signal and the internal power supply of the system. Since the quality of the power supply is influenced to a great degree by the signals passing through the rectifier it is necessary to examine in detail how system specifications affect the dimensions and performance of the rectifying elements.

Figure 4.14 shows the circuit for the full-wave bridge rectifier with models for the main parasitic elements. An internal shunt supply rectifier is not included in the model, since it has a negligible impact on the supply impedance level at high frequencies (assuming a low-power active rectifier as shown in section 4.1).

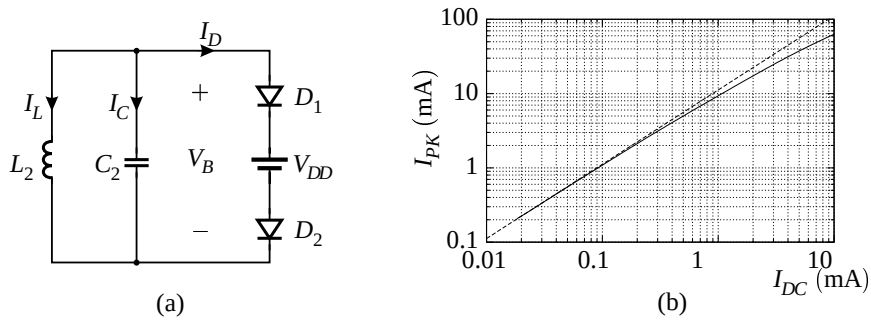


Figure 4.15: *(a) A simplified model of the bridge rectifier during one half-cycle. (b) The ratio between the peak rectifier current and the time-averaged (DC) current for the full model in figure 4.14, shown with a constant-ratio asymptote.*

Rectifier transients

The peak transient current through the rectifiers is an important characteristic of the power conversion system. Rectifiers with a large area allow the current to pass through with a lower voltage drop, and therefore a lower power dissipation, but increase the nonlinear load seen by the link, and obviously increase the die area. Another reason to quantify the peak current will be given below, where it is shown that parasitic substrate currents are strongly dependent on the current density through CMOS rectifiers.

Before the system in figure 4.14 can be subjected to analysis, it must be simplified to some degree. We choose to eliminate all parasitic elements, and assume that the system supply voltage beyond the bridge is a stiff voltage source as shown in figure 4.15. It is then possible to write a set of equations describing the current through the three main elements as $\frac{dI_L}{dt} = V_B/L$, $I_C = C\frac{dV_B}{dt}$ and

$$I_D = I_S \exp\left(\frac{V_B - V_{DD}}{2U_T}\right) \tag{4.38}$$

These lead to the nonlinear differential equation

$$\frac{d^2V_B}{dt^2} + \frac{I_S}{C}\left(\frac{V_B - V_{DD}}{4U_T^2}\right)\exp\left(\frac{V_B - V_{DD}}{2U_T}\right)\frac{dV_B}{dt} + \frac{V_B}{LC} = 0 \tag{4.39}$$

This equation can be reduced to a set of first-order equations by using the definition $u = \frac{dV_B}{dt}$,

$$\begin{aligned} \frac{du}{dt} &= -\frac{I_S}{C}\left(\frac{V_B - V_{DD}}{4U_T^2}\right)\exp\left(\frac{V_B - V_{DD}}{2U_T}\right)u - \frac{V_B}{LC} \\ \frac{dV_B}{dt} &= u. \end{aligned} \tag{4.40}$$

Solutions to this equation can only be found by resorting to numerical methods. A fourth-order Runge-Kutta method [45, 24, 80] was used to obtain a set of trajectories in the system's phase space. The initial conditions were taken as the voltage V_B and the current though the

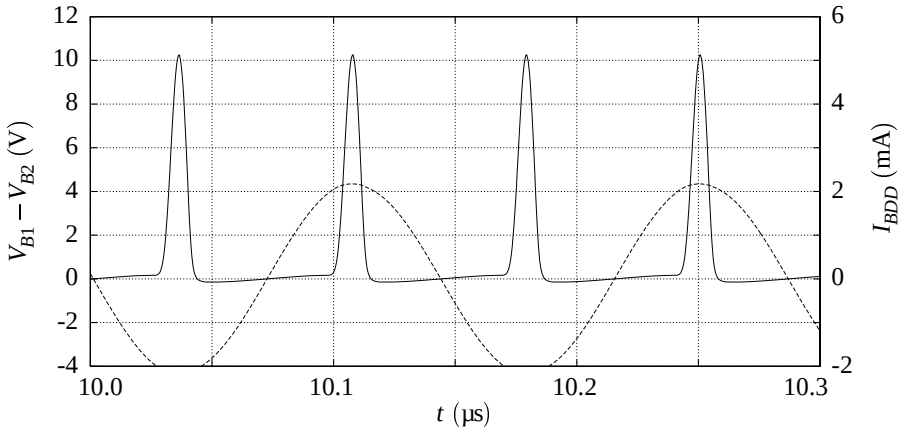


Figure 4.16: *The differential input voltage to the bridge rectifier and the resulting current through the bridge in figure 4.14 at a DC supply level of 500 µA.*

inductor a time zero. The solutions give both the peak rectifier current and the average supply current through the bridge. The numerical solutions obtained by solving (4.39) do however not take into account several effects that are of some importance. Besides ignoring the parasitic elements, chief of which is the loss resistor in the resonant circuit, the assumption of a constant supply voltage leads to an overestimate of the peak current. In addition, the model is not coupled to external fields, so that no energy is supplied to the system during the current pulse. Because of the limitations of this method, the results will not be discussed further. Instead, figure 4.15 show the results obtained by applying a circuit simulator (Spectre) to the full circuit in figure 4.14 coupled to a transmitter.

The fact that the peak current is a nearly constant multiple (10–12 times) of the average current can be explained phenomenologically by observing that the conduction angle of the bridge is nearly constant at approximately $\theta_c = \pi/5$ per half-period, independently of the load. Since the current pulse is roughly triangular in shape, the average current is $I_{DC} = \theta_c I_{PK}/(2\pi) \approx I_{PK}/10$.

This kind of hand-waving argument may fail to convince, but as a guideline it is sufficient. In a specific power extraction circuit, the peak rectifier current can be found by simulation of the complete circuit. The simple ratio will be used in the following as a rule of thumb to extrapolate the current-handling requirements of a general power conversion circuit from the system's supply current.

Full bridge current spectrum

The power spectral density of the current that passes through the bridge into the internal supply voltage node V_{DD} plays an important role in determining the total power supply noise, since it generally is the origin of the largest current transients (unless digital circuits are run directly from V_{DD} without buffering). Knowledge of the input current PSD and the impedance of the supply permits the prediction of the supply noise PSD, in the absence of load transients. Since many analog loads used in implanted systems are quite well-behaved,

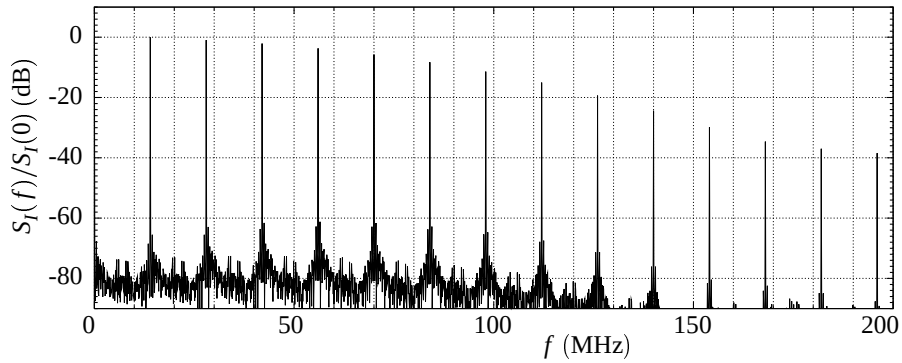


Figure 4.17: *The power spectral density for the current passing through the bridge rectifier into the V_{DD} node. The DC current is 200 µA, and the PSD is normalized with respect to the DC power. The carrier frequency is 7 MHz.*

the main part of the supply noise may come from the power link.

Because of the doubling effect of the bridge, there is no trace of the carrier frequency in the output (assuming perfectly matched elements). Instead, spectral components appear at all even multiples of the carrier frequency. For a more quantitative measure of the PSD, we observe that the time-domain signal is a train of Dirac delta functions $K(t) = \sum k_n \delta(t - nT/2)$ folded with the time-domain function $I_1(t)$ of a single current pulse. The spectrum of the pulse train is then the spectrum of the single pulse multiplied by the spectrum of the delta functions, which is

$$K(\omega) = \sum_n \delta(\omega - \frac{4n\pi}{T}) \tag{4.41}$$

where $T = 2\pi/\omega_c$ is the carrier period. The exact form of the single pulse spectrum shape depends on the pulse, but assuming a perfect Gaussian pulse shape, the respective $1/e$-amplitude pulse widths in the time and frequency domain are complementary, and their product is 2 [49, 23]. This relation is not exact for non-Gaussian pulses, but it is still correct within a factor of 2 for well-behaved pulse functions. Using half the conduction angle of the bridge rectifiers for the current pulse, the $1/e$-amplitude width b_ω of the pulse spectrum is approximately

$$b_\omega = \frac{2}{b_t} = \frac{2\pi}{\theta_c T} = \frac{\omega_c}{\theta_c} \tag{4.42}$$

Figure 4.17 shows the power spectral density for a simulation run at a DC supply current of 200 µA. As the expression for the delta functions' spectrum indicates, the first few even harmonics have an amplitude very close to the DC level, and their amplitude falls off for frequencies higher than the bandwidth of the single pulse. The "noise floor" on the graph is a numerical artifact.

Because of the nonlinear response, the spectrum obtained for one DC current level cannot be scaled directly to other current levels. To examine how the spectrum depends on the DC supply current, a large number of transient simulations were performed with the DC current ranging from 10 µA to 10 mA, spanning the range of most low-power implanted devices. Their spectra were calculated using a discrete Fourier transform and the power at

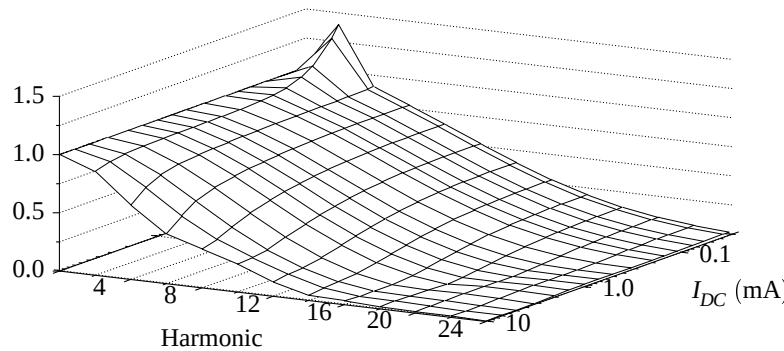


Figure 4.18: *The relative amplitudes (not power) of the bridge current harmonics as a function of DC system supply current.*

each even harmonic was then integrated numerically. The resulting distributions are shown in figure 4.18, with the amplitude of each harmonic normalized with respect to the DC value. The amplitude is used instead of the power, both because there is no natural normalizing impedance level, and because amplitudes are usually more useful in discussions of power supply noise. The figure only shows the even harmonics of the carrier frequency as the odd harmonics have zero amplitude in a matched bridge rectifier.

The figure has several interesting features which bear mentioning. The spectral content at higher frequencies decreases as the DC current level is increased, which fits the previous results which showed that the current pulses become proportionally wider as the amplitude increases. Another feature is that the amplitude at twice the carrier frequency is increased in relation to the other frequencies as the DC current falls to very low levels. This phenomenon is due to the capacitive feed-through in the rectifiers. The first non-zero harmonic is typically the most important in a system with a capacitive power supply impedance, so this can be an important effect in very low-power systems.

Signal distortion

The nonlinearity of the load seen by the antenna can cause a significant amount of distortion in the signal if a shared channel is used for data and power. In the typical setup with power and data transmitted over the same link, this distortion should be accounted for. The distortion characteristics must inevitably be allowed to influence the choice of modulation method and the applied power levels.

The first apparent effect of the use of rectifiers for power conversion is a soft limiting of the signal. This leads to a compression of the transfer characteristic with the resulting harmonic distortion and intermodulation.

As an aside, it is interesting to compare this problem to a similar problem encountered in power electronics, namely the nonlinear load that most simple AC–DC converters present to the 50/60 Hz distribution net. The problem can be solved at low frequencies by the use of ac-

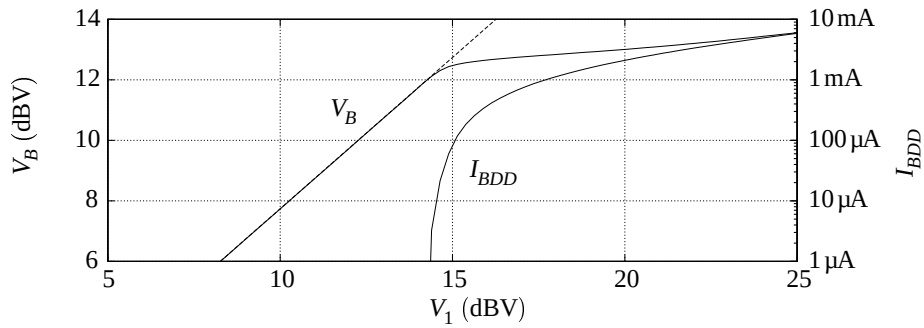


Figure 4.19: *The fundamental bridge input amplitude as a function of the primary amplitude. The input reference level is arbitrary because of the variable channel gain. The output is referred to a 1 V level. The compression point occurs at the point where the rectifier elements start to be forward-biased. The right axis shows the time-averaged current through the bridge.*

tive power factor correction circuits [14], which replace the basic rectifier configuration with a switching regulator network. The method would however be difficult to implement for load linearization in implanted system power circuits, since it demands switching frequencies that are far higher than the line frequency, and the use of inductive elements.

The fact that the signal compression limit is dependent on the internal supply voltage of the system complicates matters since the response is effectively time-dependent, and dependent on the past history of the input signal. By assuming that the time scale for the changes in the system is much longer than the changes in the signal, the system can be considered time-independent to first order.

In order to apply the usual nonlinearity measures it is useful to define the channel gain with the input point defined as the voltage V_1 over the primary (external) inductor, and the output as the differential input voltage V_B to the bridge rectifier. Since the system is weakly coupled, we assume that the primary voltage is not affected by the nonlinearity. The coupling between the two parts can vary by a large factor, and the channel gain therefore also. Normal operation of the system implies a certain voltage induced on the secondary, or a certain power delivered to the load, so it is appropriate to refer all performance measures for the channel to the output instead of the input.

Moderately nonlinear channels or amplifiers are usually treated by expanding their transfer characteristics in a power series [63, 83],

$$V_B(V_1) = c_0 + c_1 V_1 + c_2 V_1^2 + c_3 V_1^3 + \ldots \tag{4.43}$$

The *n*th-order harmonic amplitude then grows as the *n*th power of the input amplitude, and a general linearity measure is the point at which the third-order term amplitude equals the fundamental amplitude (the third order intercept, IP3). In our case, the nonlinearity is so severe that the third-order term is not sufficient to describe the behavior.

Figure 4.19 shows the amplitude of the fundamental as a function of the input amplitude. This is a simulation result obtained using the same circuit definition as before, from figure 4.14. The compression point is evident, and it occurs at the point where the rectifier bridge begins to pass current into the internal supply.

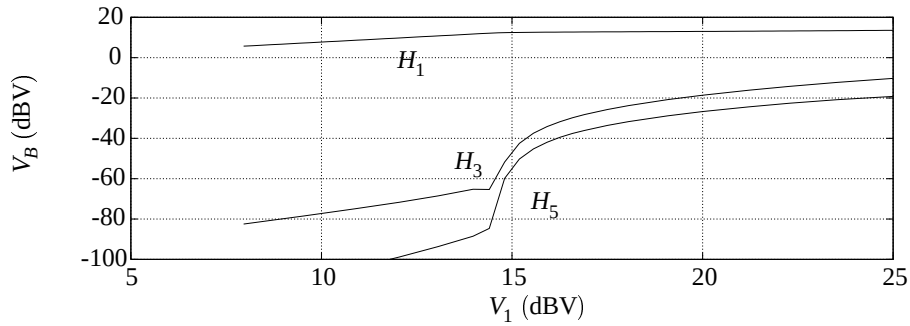


Figure 4.20: *The amplitude of the fundamental frequency and the third and fifth harmonics at the bridge input terminals as a function of the primary amplitude.*

The first two nonzero harmonics of the output signal are shown in figure 4.20. These are the odd harmonics since the symmetry of the stage eliminates the even-order terms. Instead of the roughly straight lines that would be seen were relation (4.43) obeyed, the harmonic amplitudes change drastically at the compression point. Below this point, the amplitudes vary approximately as the nth power of the input amplitude, and this is probably due to the nonlinear junction capacitances of the rectifiers. In any case, their amplitudes are 80 dB or more below the carrier, so that they have no practical importance. Close to the compression point and above it, the behavior of the harmonics amplitudes cannot be explained by the simple distortion model in (4.43).

The influence of the nonlinearity on the available modulation method and power operation region can be considered in two ways, by regarding the behavior in the time and frequency domains. Starting with the frequency domain, we see that the odd terms in the output can lead to difference intermodulation products that lie inside the signal band. Modulation methods like simple AM or ASK, which are frequently used for communication with implants [58, 4, 66, 93, 94, 98], have an input signal which contains different spectral components that are susceptible to intermodulation. The modulation form and modulation index must therefore be chosen so that for the target operating point of the rectifiers, the intermodulation terms can be handled by the receiver.

Considering the nonlinearity in the time domain, specifically with reference to figure 4.19, we see that the AM modulation of the input signal is suppressed as the output moves past the compression point. The minimum operating distance along the V_1 axis in that figure is set by the power requirement of the system, so that in the AM case at least, higher power levels require a higher input modulation index for a given receiver sensitivity.

As a last observation on the nonlinearity induced in the input signal by the rectifier network, we note that the link should always be operated at the minimum possible power. The excess supply power can be dissipated in an internal shunt, but the higher the power level for a given system, the higher the signal distortion. This minimum power requirement is of course compatible with the desire to obtain maximum battery life in a portable system, but may be difficult to implement in practice. An external transmitter is normally operated at a power level exceeding the level required by the implant, to ensure that the required power is transferred even under a worst-case misalignment. This is another reason to include adaptive

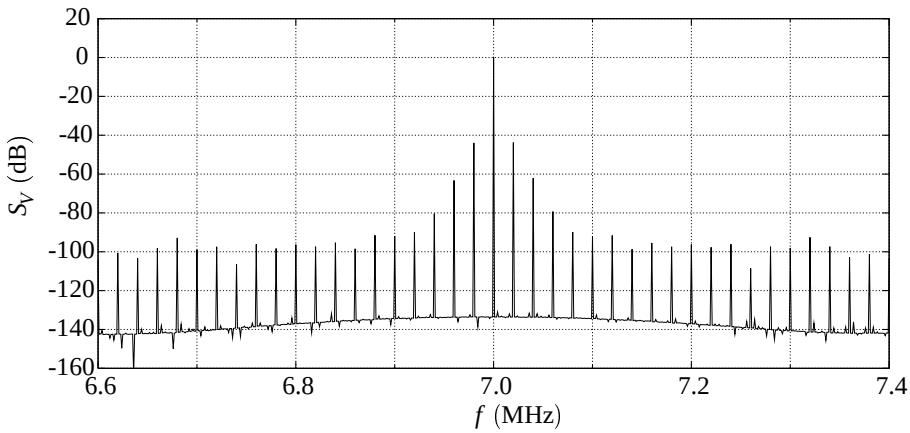


Figure 4.21: *The power spectrum of the signal at the bridge input terminals, referred to the fundamental term. The primary driving signal was AM modulated with a modulation index $m = 0.1$ and a modulation frequency of 100 kHz. The ± 100 kHz sidebands, which would have been at -40 dB in a linear channel, are suppressed, and there are intermodulation terms at all multiples of the offset frequency.*

power transfer with sensing of the internal supply voltage.

4.2.3 CMOS full bridge

There are several ways to implement rectifying elements in CMOS, each with its own quirks. Diode-connected MOS transistors of both polarities can be used, with the associated square-law current characteristic (at least if they are dimensioned for operation in strong inversion). Unless the CMOS process in question features twin wells, one type of transistor is going to be more attractive than the other. A rectifier transistor residing directly in the substrate (most likely an NMOS transistor nowadays) will cause a lot more substrate noise than one sitting inside a well. For this reason, a solution relying on MOS transistors for rectification should use only p-channel devices if possible. The drawback is that p-channel transistors usually have 2–4 times less current gain compared to n-channel transistors with an identical area and parasitic capacitances.

Another possibility is to use different diffusions to implement true p-n junction diodes. Again considering the case of an n-well process, the possible combinations are n+/substrate, n-well/substrate and p+/n-well. The first two have the substrate as one terminal, and as such are unsuitable because of the substrate coupling. The last possibility involves using a p+ diffusion area inside an n-well as the anode and the well as the cathode. This poses a problem in the case of the V_{SS} diode (with the anode connected to V_{SS}) in a bridge rectifier. The input node is the well, and during normal operation its potential falls below V_{SS}. A current can therefore flow directly from the p- substrate into the well, defeating the purpose of using n-well junctions. The magnitude of the unwanted current depends on the respective characteristics of the p+/n-well and p-/n-well junctions and their areas. In order to maximize the ratio of the current through the p+/n-well junction to the unwanted current, the area ratio

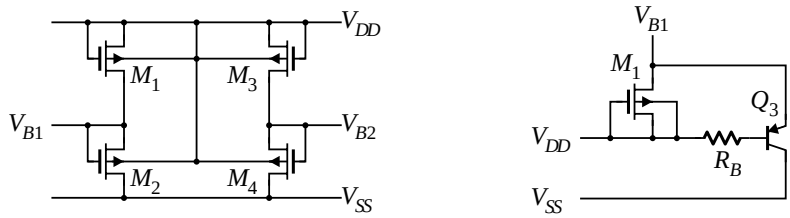


Figure 4.22: *A full-bridge rectifier implemented using diode-connected PMOS transistors in a common well. The detail on the right shows the vertical PNP associated with the input V_{B1}. An identical transistor is connected to the second input.*

must be as high as possible. The area ratio will be around 1 in a reasonable layout (values slightly larger than 1 may be possible by exploiting sidewall areas in a waffle-iron layout). The process parameters quote reverse-bias leakage currents for the junctions in question, and keeping in mind the diode current expression $I_F = I_S(\exp(V_F/U_T) - 1)$, this current is seen to be identical with the scale current I_S. The ratio of the scale currents for the two junction types is only 1–2, so a considerable part of the diode current would pass by the alternate path through the substrate if such a rectifier junction were used.

Another reason not to use p-n junctions in the rectifier implementation is the presence of parasitic bipolar transistors. If a p+/n-well junction is used to implement the V_{DD}-connected rectifier in figure 4.14, the well must be connected to V_{DD}. The p-n junction is then the emitter-base junction of a parasitic PNP whose collector is the substrate. A fraction $\beta/(\beta+1)$ of the input current will then flow directly to V_{SS} instead of to V_{DD}. It is possible to reduce the current gain β of the parasitic transistor by using clever layout tricks, but a large fraction of the input current will still be lost.

In light of these results, it is apparent that the CMOS structures discussed above are not suitable for power supply rectification. Solutions using PMOS transistors were therefore examined in more detail. An implementation using PMOS transistors for on-chip rectification on smart cards has previously been presented in [44], but the question regarding the elimination of substrate current was not addressed. A full bridge rectifier using PMOS transistors in a common well is shown in figure 4.22. The figure does not show all parasitic elements which are important for the function of the circuit, but an idealized cross section through one of the rectifier branches in figure 4.23 shows the circuit in more detail. The presence of the parasitic transistor Q_3 is the main problem. During normal operation of the rectifier, its emitter-base junction becomes periodically forward-biased, and current can flow from the input point V_B directly to V_{SS}. (The transistor Q_3 is shown with two components, but M1 and M2 do of course share a common diffusion in a real layout). The question is then whether the MOS transistors can be designed so that a negligible portion of the total current is lost in the parasitics. To answer that, it is instructive to compare the current through the MOS and bipolar transistors. At low current levels, the currents through both the MOS and bipolar devices are exponential functions of the applied voltage $V_{SD} = V_B - V_{DD}$. The collector current of the bipolar transistor is

$$I_C = I_S \exp\left(\frac{V_{SD}}{U_T}\right) \tag{4.44}$$

The drain current of the MOS is rather more complex, depending among other things on the

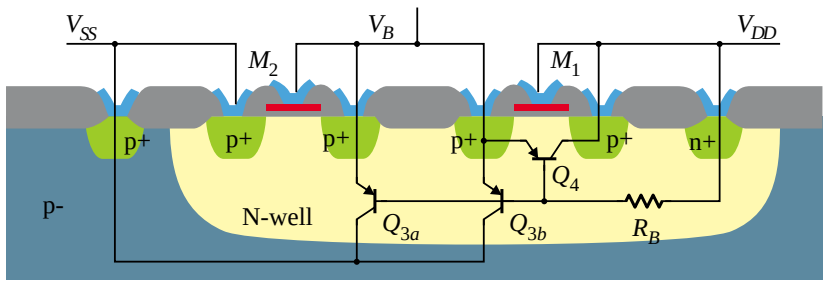


Figure 4.23: *An idealized cross-section through one CMOS bridge rectifier branch, showing parasitic bipolars.*

inversion condition and the connection of the bulk node. The subthreshold slope factor n which appears in the current expression for the MOS usually means that the MOS current increases slower than in a bipolar transistor with the same bias current. The respective currents would then have an intersection point when plotted as a function of the applied bias, above which point the bipolar current would dominate. Regarding the bridge circuit in figure 4.22 specifically, we will show that this is not the case, at least over a portion of the operating range. There is a range of bias currents where the ratio between the currents through the MOS and the bipolar transistor is constant. This is one of the features of the circuit which makes it attractive for use as a rectifier.

The lateral parasitic PNP, marked as Q4 in the figure, also contributes to the input current. This transistor is however connected in parallel with the main transistor M1, and does not introduce any undesirable parasitic effects. It does affect the behavior of the rectifier to some degree, by contributing a current which is proportional to $\exp(V_{SD1}/U_T)$ to the input current. This contribution is however indistinguishable from the current through M1, as long as M1 is biased in its target regime.

Full operating range behavior

It is possible to derive conditions on the device characteristics which ensure that the fraction of the current which is "lost" through the bipolar transistor is below a certain limit. Those conditions will be derived here starting with general operating principles, and thereafter connecting the results to process and layout parameters, in order to obtain useful design equations.

We have already mentioned the fact that the usual exponential expression for the MOS drain current in weak inversion contains a $1/n$ factor in the exponent. In making that assertion, we ignored the fact that the bulk of transistor M1 is not connected to the source. In fact, the bulk is connected to the drain, a situation which is rather unusual, and only rarely encountered in analog CMOS circuits.

To obtain an accurate value for the drain current, it is necessary to take the variation of the source-bulk voltage into account, since this voltage is equal to the source-drain voltage. The traditional methods would be to include the bulk effect through a threshold voltage which is a function of a bulk factor γ and the source-bulk voltage, among other things. A more elegant method is to use the full EKV model for a transistor in weak inversion [34], since

the model already uses terminal voltages which are referred to the bulk. According to this model, the drain current expression for a PMOS transistor is given by

$$I_D = I_{D0} e^{-V_G/nU_T} \left[e^{V_S/U_T} - e^{V_D/U_T} \right] \tag{4.45}$$

with the scale current I_{D0} defined by

$$I_{D0} = 2n\mu C_{ox} \frac{W}{L} U_T^2 \exp\left(\frac{V_{T0}}{nU_T} \right) \tag{4.46}$$

The threshold voltage inside the exponential can lead to large variations in the scale current; a worst-case process variation in the threshold voltage of 150 mV away from the nominal value, leads to a scale current that is off by a factor of 50.

The drain, gate and bulk of the top transistors in the bridge rectifier (M1 and M3 in figure 4.22) are connected to the supply voltage V_{DD}. The current through the top transistors can therefore, when they are forward-biased, be expressed as a function of the forward voltage drop $V_B - V_{DD}$,

$$I_D = I_{D0} \exp\left(\frac{V_B - V_{DD}}{U_T} \right) \tag{4.47}$$

This is very interesting since is has the same form as the dependence of the collector current in a bipolar transistor on the base-emitter voltage (4.44). The only difference between the drain current of M1 in weak inversion and the collector current of the parasitic PNP is therefore a constant factor. This can be written as

$$\frac{I_{D,wi}}{I_C} = \frac{I_{D0} e^{V_{SD}/U_T}}{I_S e^{V_{EB}/U_T}} = \frac{I_{D0}}{I_S} \tag{4.48}$$

which is independent of the applied bias, at least in the weak inversion limit; the ratio depends only on process parameters and the geometry. The exact numerical value of this ratio tells us how effective the bridge rectifier construction is in reducing the parasitic bypass current to the ground node. In the following sections, we shall examine the dependence of this ratio on geometry and technology scaling.

Another advantage of using the rectifier transistors in weak inversion, besides the reduction of parasitic currents, is the reduction in power dissipation which follows from the reduced voltage drop. If the top transistors are dimensioned for operation in weak inversion, and the absolute value of the threshold voltage is 0.6 V (not unusual for modern-day processes), the power dissipation in the top transistors in figure 4.22 may actually be lower than in corresponding junction diodes.

There is another side to the power dissipation question. The bottom rectifier transistors in the bridge do not need to operate in weak inversion, at least not for the same reason as the top transistors, since they are not plagued by the same parasitic bipolars. They do however suffer from a large source-bulk bias, which increases their threshold voltage and power dissipation. In fact, the source-bulk bias is as large as it can possibly be in a single-supply circuit, since it is equal to the supply voltage. The power penalty due to the bulk effect can be reduced by biasing the transistors in weak inversion, and in that way reducing the voltage drop over them. Assuming that (4.45) is applicable, the influence of the bulk effect is equivalent to using a threshold voltage

$$V_T = V_{T0} - (n-1)V_{DD} \tag{4.49}$$

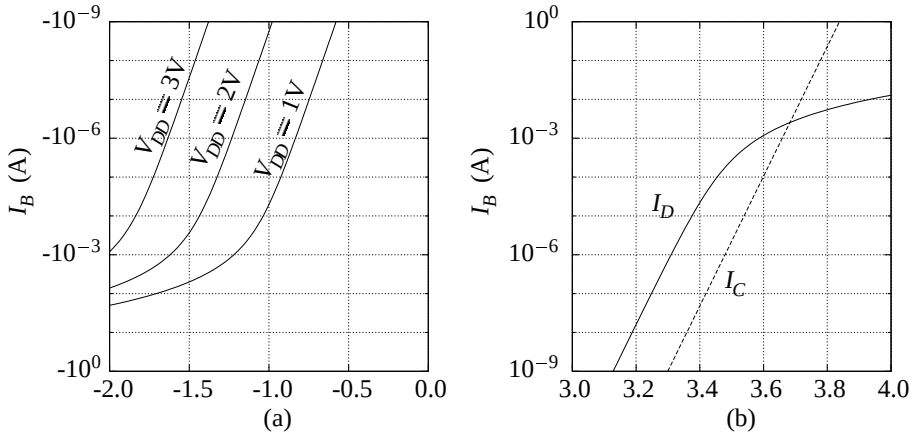


Figure 4.24: *The input current to the PMOS bridge rectifier as a function of the input voltage V_B, calculated according to the EKV model from weak to strong inversion. Figure (a) shows the current through the V_{SS}-connected bottom rectifier, as the input voltage V_B falls below $V_{SS} = 0\,V$. Note the unusual logarithmic scale. The input current was calculated with three different values of the supply voltage V_{DD}, to illustrate the influence of the bulk connection. Figure (b) shows the current through the V_{DD}-connected rectifier when the input voltage exceeds the supply voltage of 3 V. The figure also shows the current through the parasitic bipolar transistor, modeled by* (4.44) *with a scale current $I_S = 10^{-14}A$.*

This result shows that the shift in threshold voltage is proportional to the source-bulk bias, in contrast to the strong-inversion bulk effect normally described using the bulk effect constant γ [55, 59]. The change in threshold voltage can obviously be quite large, typically on the order of 1 V for a 3 V supply voltage, so it is worthwhile to reduce the voltage drop over the bottom rectifiers as much as possible by using wide transistors and weak inversion.

The characteristics of the V_{DD}-connected rectifiers M1 and M3 should be completely independent of the supply voltage, since none of its four terminals is connected to V_{SS}. There is however a related secondary effect, which causes a slight increase in the input current with the supply voltage, all other things being equal. It is due to the parasitic PNP, which has its collector connected to V_{SS}. This is only of minor importance, since we are proceeding with the assumption that the share of the PNP in the total current is small.

The input current to the bridge rectifier was calculated using the EKV MOS model, and the results are shown in figure 4.24. They apply to an implementation of the circuit in figure 4.22, with all transistor dimensions set at 800 µm/0.5 µm, and process parameters for a standard CMOS process. The calculation is valid all the way from weak to strong inversion, since a relatively accurate interpolation function was used in the moderate inversion region [34]. The transistor dimensions coincide with those used in the test implementation of the circuit.

The calculation results agree with the qualitative results of the discussion above. The voltage drop over the V_{SS}-connected rectifiers increases linearly with the supply voltage, reaching the relatively high level of 2 V for a 3 V supply voltage and a 1 mA bias current. The voltage drop increases rapidly outside the weak inversion region, which is another reason to

design the circuit for operation in that regime. However, once the transistors are made wide enough to attain weak inversion, we gain very little by using wider transistors. This point of diminishing returns exists because the width (and the area) must be increased by a factor e to reduce the voltage drop by $nU_T \approx 40\,\text{mV}$.

The input current corresponding to the positive half-cycle of the input voltage, when the input voltage rises above the supply voltage, is shown in the second half of the figure. The two components of the current are shown, through the MOS and through the parasitic PNP. The MOS current was again modeled using an interpolation function for the moderate inversion region. The bipolar current was modeled using the simple exponential model of the transistor (4.44) and a scale current corresponding approximately to the expected value. The bipolar model did not include the base resistance, which is considerable in this type of vertical PNP, and is influential at high collector current levels. As expected, the drain current is larger than the parasitic current by a constant factor while the transistor is in weak inversion. The dependence of this factor on the layout geometry and the process parameters is the subject of the following section.

Geometry effects

It has already been established that a certain fraction of the current though the top rectifier transistors in figure 4.22 is lost through the parasitic bipolar transistors. It has also been shown that this fraction is independent of the operating point as long as the transistors are in weak inversion, with the value of the ratio given by (4.48). This ratio is formally very simple, but it includes a significant number of free parameters. Obviously, to obtain the best performance from the bridge rectifier it is best to design a layout which maximizes the MOS scale current I_{D0} in relation to the bipolar scale current I_S.

The bipolar scale current I_S is the product of the saturation current density and the effective emitter area, $I_S = J_S A_e$. The current density J_S depends on intrinsic carrier electron density, dopant concentration and carrier diffusion constants [40]. These are not usually known quantities, especially since they are used here to model a parasitic device. Instead of applying an analytic description to the transistor, we must therefore attempt to extract useful information from the available data. Parasitic bipolar transistors are in fact modeled sometimes in analog CMOS processes since they can be useful in some circuits. The approach here will be to use an empirical value for the scale current density, since it appears to be relatively constant between different technologies. The average scale current at room temperature (300K) for vertical parasitic PNP transistors in several submicron technologies was found to be $4.5\times10^{-19}\ \text{A}/\mu\text{m}^2$, with the highest value only two times larger than the lowest one.

The geometry of the rectifier enters into the design through the effective emitter area of the PNP transistor and the aspect ratio W/L of the MOS transistor. The equations can be simplified by defining a geometry-independent MOS scale current I_M through $I_{D0} = I_M A_M$ where A_M is the MOS aspect ratio. Referring to (4.46), we see that

$$I_M = 2n\mu C_{ox} U_T^2 \exp\left(\frac{V_{T0}}{nU_T}\right) \tag{4.50}$$

Using these definitions, we can express the ratio of the MOS current to the parasitic current

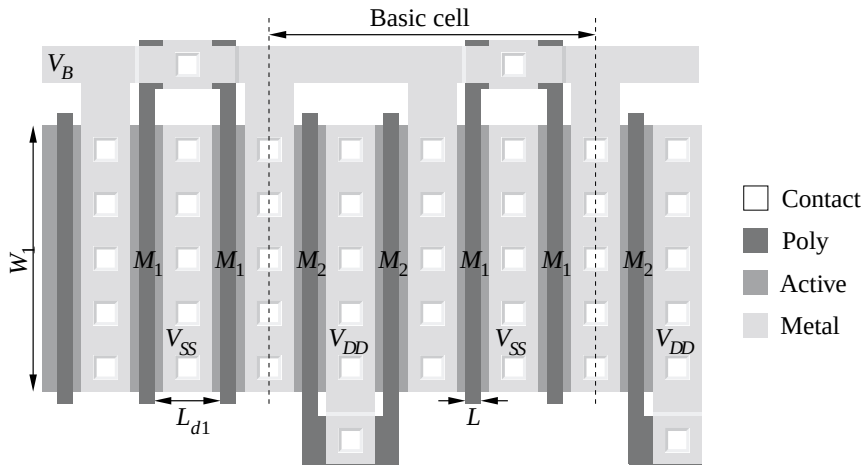


Figure 4.25: *A fingered layout example for the two transistors in a rectifier branch. Not drawn to scale.*

as

$$\frac{I_{D0}}{I_S} = \frac{I_M}{J_S}\frac{A_M}{A_e} \tag{4.51}$$

The first factor depends only on process parameters, while the second is, in this simple model at least, purely geometrical. It is true that the diffusion depth and other processing details do have an influence on the effective emitter area and the MOS geometry, but for the purposes of this derivation, such a separation of variables is helpful.

The task is then to design the optimal layout which gives the highest value of the geometry factor, which we define as $\Xi = A_M/A_e$. The effective emitter area for a given diffusion depends on both area and sidewall terms. The importance of the sidewall term can be estimated from figures for diffusion capacitances, but this estimate will be inexact since different mechanisms are at work in defining capacitances and effective emitter areas. In the following derivations, the contribution of the sidewall terms to the emitter area will be ignored.

Figure 4.25 shows a layout example which places both MOS transistors in a rectifier branch in a single fingered structure. The MOS transistors should of course always be drawn with minimum gate length, to increase Ξ. The only reason not to do so would be because of yield considerations. The figure shows a scaled-down version of the layout for clarity, but the fingers in a practical layout are so long that edge effects are negligible. At the low frequencies in question and with salicided gate polysilicon, the gate resistance of the bridge transistors is negligible, even for relatively long gate poly strips ($W_1/L \approx 100$). The emitter area for the basic cell in this example is $A_e = 2W_1L_{d1}$, and the aspect ratio of M1 in a single cell is $A_M = 2W_1/L$. The geometry factor in this case is then

$$\Xi_1 = \frac{1}{LL_{d1}}. \tag{4.52}$$

Not surprisingly, shorter gate and diffusion lengths improve the performance of the device.

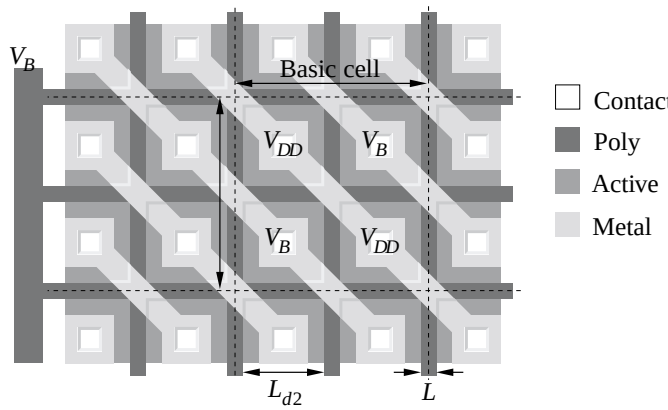


Figure 4.26: *A waffle-iron layout for a rectifier transistor. Not drawn to scale.*

This example placed both transistors in the bridge branch in the same structure, but that is not necessarily the best solution. Another possibility is to lay M1 and M2 out separately. This splitting of the layouts offers the possibility of using different widths for the two transistors. Using a wider M2 could in some circumstances be desirable, since the voltage drop over M2 is higher than over M1 because of bulk modulation of the threshold voltage. The Ξ factor is not improved by using a wider M2 however, since the effective emitter area is still proportional to the sum of the two widths, while the aspect ratio of M1 is reduced.

An alternative waffle-iron layout for M1 is shown in figure 4.26. Because the diffusion areas must be contacted in some way, it is usually not possible to use minimum dimensions for the drain and source diffusions. For the layout shown in the figure, the widths of the diffusion areas must be increased to accommodate the diagonal metal strips. In the previous example, the width was set by the contact dimensions, and contact-to-gate spacing, but metal width and spacing rules also play their part here. The effective emitter area for the basic cell shown in the figure is $A_e = 2L_{d2}^2$, but assuming that M2 is laid out with equal dimensions to M1, the area must be doubled to $A_e = 4L_{d2}^2$. The effective MOS gate length is near L and the gate width is approximately $W = 8(L_{d2} + L)$. The geometry factor is then

$$\Xi_2 = \frac{2(L_{d2} + L)}{LL_{d2}^2}. \tag{4.53}$$

The waffle layout is therefore superior to the finger layout by a factor

$$\frac{\Xi_2}{\Xi_1} = \frac{2L_{d1}(L_{d2} + L)}{L_{d2}^2}, \tag{4.54}$$

which is close to 2. The analysis did not include periphery effects for the emitter area, which are presumably more important in the waffle case. The area advantage in using a waffle layout is therefore not quite as great as (4.54) indicates. The potential gain is still large enough to justify the increased layout complexity.

The effective length and width of the waffle transistor layout was found by a rather crude approximation. The subject is treated in more detail in [43], which gives a treatment of

effective geometries of enclosed-gate layout transistors (transistors with drain island within a gate ring, which is in turn enclosed by the source). Another reference [12] shows that waffle transistors give better matching than traditional finger structures in modern processes, where the mismatch is dominated more by piezoelectric strain effects in the die than by process gradients.

Conversion efficiency

The conversion efficiency of the CMOS bridge converter is lower than for the junction diode version because the sum of the rectifier voltage drops is higher. When calculating the efficiency of a bridge rectifier, the voltage drop over both the top and bottom rectifiers in a branch must be summed, and as we have seen figure 4.24, the drop over the bottom PMOS rectifier is quite large.

An exact treatment of the power dissipation in the converter requires the integration of the current-voltage product over time, with all parasitics taken into account. To obtain an explicit expression for the power dissipation and the efficiency, we may however make a few simplifying assumptions about the system, and still obtain a result which is correct to within a factor of 2 or so. The near-constant relationship between the peak rectifier current and the DC supply current of the system has already been pointed out, and the ratio shall be denoted here by the constant χ ($\chi \approx 10$ for a bridge converter). The exponential dependence of the current on the rectifier bias leads to the second simplification, namely the peak forward drop approximation (4.36). The power dissipation is then

$$P_d = I_{DC} V_{drop}(I_{peak}) = I_{DC} V_{drop}(\chi I_{DC}) \tag{4.55}$$

The total voltage drop is the sum of the source-drain voltages over M1 and M2 in figure 4.22. The transistors are presumably operating in weak inversion, so the two components of the voltage drop are found from (4.45) and (4.47) as

$$V_{SD1}(I_D) = U_T \ln \frac{I_D}{I_{D0}} \tag{4.56}$$

and

$$V_{SD2}(I_D) = nU_T \ln \frac{I_D}{I'_{D0}}, \tag{4.57}$$

where the modified scale current I'_{D0} includes the effect of the bulk bias of M2,

$$I'_{D0} = 2n\mu C_{ox} \frac{W}{L} U_T^2 \exp\left(\frac{-V_{T0}}{nU_T}\right) \exp\left(\frac{(n-1)V_{DD}}{nU_T}\right). \tag{4.58}$$

The total voltage drop is then

$$V_{drop}(I_D) = (n+1)U_T \ln I_D - U_T \left(\ln I_{D0} + n \ln I'_{D0}\right). \tag{4.59}$$

This expression might not be quite correct formally, since it involves logarithms of quantities which are not dimensionless. It can however be rewritten as the logarithm of a dimensionless

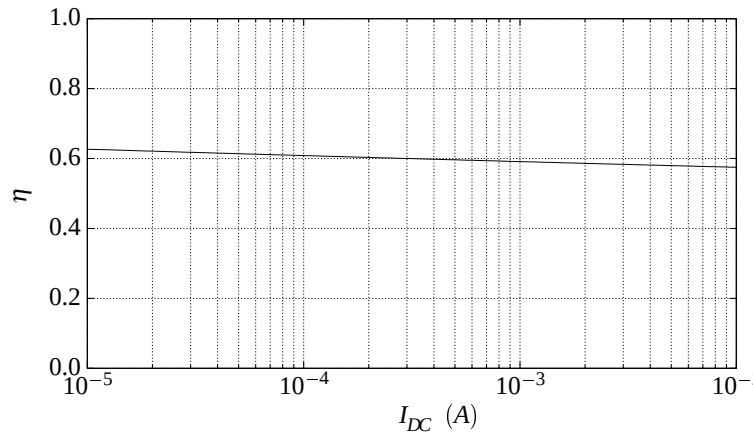


Figure 4.27: *The efficiency of the PMOS bridge power conversion circuit as a function of the DC supply current of the chip, with a supply voltage of 3 V. The dimensions of the transistors are 800/0.5, as in the test circuit. Other parameters are typical for a 0.5 µm process.*

fraction. Setting the peak current equal to χI_{DC} and using the definition of the scale currents (4.46) and (4.58), the power dissipation is

$$\begin{aligned} P_d &= I_{DC}\left[(n+1)U_T \ln(\chi I_{DC}) - 2U_T \ln\left(2n\mu C_{ox}\frac{W}{L}U_T^2\right)\right. \\ &\quad \left. -\frac{1}{n}\left((n+1)V_{T0} - (n-1)V_{DD}\right)\right]. \end{aligned} \tag{4.60}$$

Using the power dissipation of the circuits supplied through the bridge rectifier, $I_{DC}V_{DD}$, the efficiency of the power conversion can easily be calculated. The plot of the efficiency in figure 4.27 shows that it depends very little on the DC supply current, and lies in the neighborhood of 60% for a supply voltage of 3 V. The near-constant efficiency is due to the weak dependence of the forward drop over the rectifiers on the current. Resistive losses in the conversion circuit can influence this number, especially at the higher current levels, but since the DC resistance of the external and internal interconnect is on the order of 1 Ω, the error is negligible in most cases.

Implicit in these calculations is the assumption that the transistors in the bridge are in weak inversion. This condition can obviously not be met by a single implementation over an arbitrary range of operating currents. The converter is therefore only capable of functioning at the efficiency obtained from (4.60) up to a certain current level, and the efficiency drops quickly above that level as the transistors leave weak inversion. This does not really matter as long as the dropoff point lies above the normal current consumption of the chip. If the average current through the converter is higher than that used by the chip, it must be dissipated in a protection circuit like the shunt regulator in section 4.1. The underlying problem is then that too much power is transmitted into the circuit, and it does not matter whether the excess power is dissipated in the converter or in an internal protection circuit.

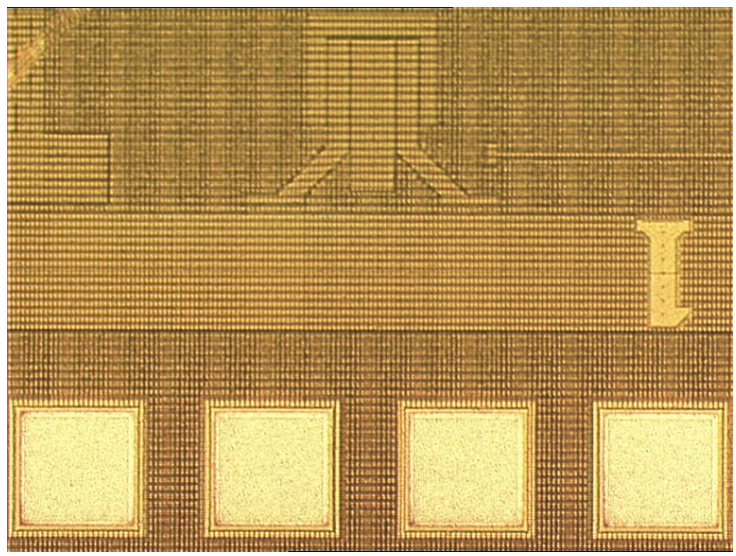


Figure 4.28: *A microphotograph of the bridge rectifier. The features of the circuit are not easily discernible because of the metal layer fill added by the wafer fab. It's the bit which looks like a Space Invader™.*

4.2.4 Experimental rectifier data

The PMOS bridge rectifier described in the previous section was implemented in a 0.5 µm CMOS process. The V_{DD} and V_{SS} terminals of the bridge were accessible externally on the test chip, as were the two input terminals V_{B1} and V_{B2}. Because the test circuit was in fact a part of a larger system, the power supply terminals were not dedicated solely to the bridge rectifier but were shared with other circuits on the same chip. The small number of available pads made it impossible to route the supply signals off-chip between the rectifier and the rest of the supply net. Such a solution would have been inferior anyway because of the parasitic elements it would have inserted in the supply path. The penalty incurred by integrating the rectifier and the load is a small decrease in the measurement accuracy because some of the measurements must be corrected for the effect of the load.

The bridge transistors were all given the same dimensions, $W = 800\,\mu\text{m}$ and minimum length $L = 0.5\,\mu\text{m}$. The transistors were laid out in an interleaved finger structure, as shown in figure 4.25 rather than a waffle-type layout which would probably have given a better reduction of the parasitic current from the input to V_{SS}. This test circuit was designed with a rather modest current consumption of 100–200 µA in mind, and the total layout area was therefore only $140 \times 100\ \mu\text{m}^2$. A microphotograph of the bridge rectifier is shown in figure 4.28.

Static measurement results

The availability of all four bridge terminals for testing makes it possible to measure the parasitic current by subtracting the current out of the V_{DD} terminal into the load from the

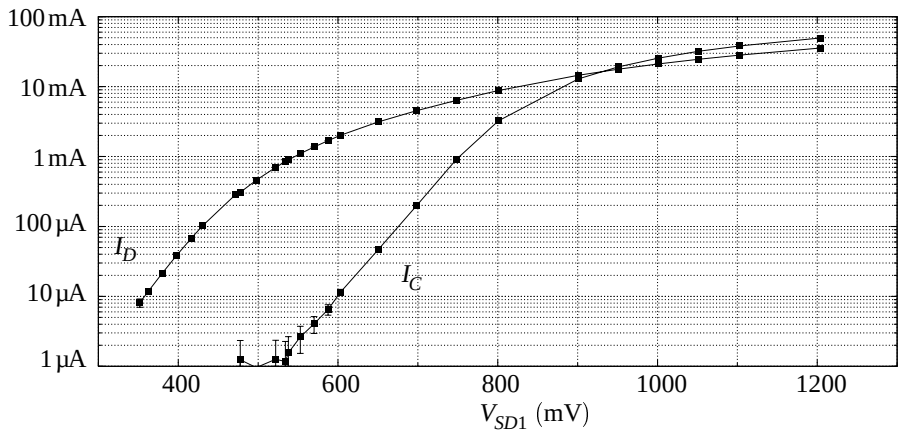


Figure 4.29: *The measured current through the rectifier transistor M1, and the parasitic PNP current.*

input current into terminal V_{B1} or V_{B2}. There were however some sensitivity problems due to the fact that some of the other circuits on the IC were connected across the power supply, and could not be disconnected for the bridge measurements. Their supply current was added to the measured load current, but combined with the fact that the small parasitic current was calculated by subtracting two large currents from each other, the limited accuracy of the internal load current measurement gave rise to considerable numerical instability in the final result. The measurement uncertainty of the parasitic currents was therefore on the order of 1 µA.

The static measurements of the experimental circuit were carried out on one branch of the rectifier only, since the two branches are independent for all practical purposes. The top transistor (connected to V_{DD}) is therefore referred to only as M1, and the bottom transistor which is connected to V_{SS} is M2.

The results for M1 are shown in figure 4.29. The figure shows the MOS and parasitic PNP currents as the input voltage is biased above the supply voltage, which was set to 3 V. At small bias voltages, with M1 in weak inversion, the slope of the MOS current is the same as for the bipolar transistor, exactly as predicted by (4.47). The ratio between the currents in this region of operation is close to $I_D/I_C = 2000$. To obtain this ratio it is however necessary to bias the transistor at very low current levels, which requires a large total area for the converter. By allowing the transistor to operate some distance out in moderate inversion, say at a source-drain voltage of 550 mV, a still respectable figure of $I_D/I_C = 300$ is obtained with a drain current of 1 mA. This peak drain current would correspond approximately to a DC supply current of 100 µA in a bridge converter.

There is an intersection point between the two curves where the parasitic current becomes equal to the MOS current. This point does however occur at a larger forward bias than the simple model used in figure 4.24(b) would indicate. The parasitic current does not continue to increase indefinitely according to an exponential characteristic, mainly because of the base resistance R_B and the finite current gain. The base resistance is governed by the sheet resistance of the n-well and by the layout of the converter, and while the exact value is difficult to calculate, a reasonable estimate is $R_B = 500\,\Omega$. The scale current of the PNP is

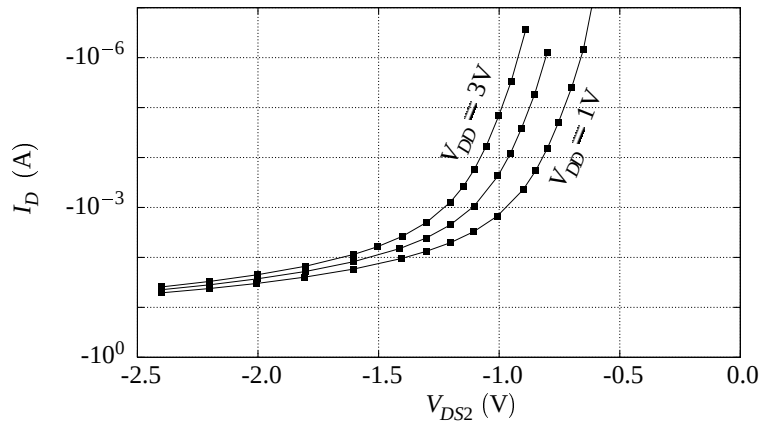


Figure 4.30: *The measured input characteristic for the V_{SS}-connected rectifier. Three sweeps are shown, with supply voltages of 1, 2 and 3 V. The influence of the bulk bias on the voltage drop is evident. Note the reversed logarithmic scale.*

easily found from the straight part of the curve in figure 4.29, and is for this particular layout equal to 3.5 fA. The base resistance and the current gain could be used as fit parameters for the measurement data, were it not for the fact that they form a degenerate parameter set in the simple model with only these two nonidealities. The equation to be fitted to the data is

$$V_{SD1} = V_B - V_{DD} = \frac{R_B I_C}{\beta} + U_T \ln\left(\frac{I_C}{I_S}\right), \tag{4.61}$$

where R_B and β are obviously not independent.

The characteristic of the bottom transistor M2 is also of interest, but for a different reason. The main drawback of using a PMOS transistor for the bottom rectifier was not the presence of parasitic currents, but the large voltage drop. The I-V characteristic of M2 was measured at three different bulk bias levels, with supply voltages of 1, 2 and 3 V. None of the circuits described here are capable of operating at 1 V, but it was included as an interesting data point, for purposes of extrapolation to other systems.

The results for M2 are shown in figure 4.30. This bridge converter was designed for peak currents of approximately 1 mA, and this current is therefore used as a reference level when assessing the voltage drop. The effect of the bulk bias is surprisingly small, and seems to correspond to a subthreshold slope factor n of only 1.2, rather than the value of 1.4 which was used in the calculations (see figure 4.24(a)). As a result, the voltage drop at the rated current and supply voltage is only 1.2–1.3 V, which compares favorably to the expected 2 V.

Although the test circuit shown here was dimensioned for a rather modest total current consumption, there is in principle nothing to prevent scaling it up to supply currents of several milliamperes. With the same fingered layout used here, and a current ratio $I_D/I_C = 300$, the ratio of total bridge area to DC supply current is 140 000 μm^2/mA. By using a waffle-iron type of layout the required area can be reduced, and this ratio approximately halved.

Dynamic measurements

The performance of the bridge rectifier was also measured under real load conditions with an inductively coupled signal source. It was however not possible in the dynamic case to separate the characteristics of the rectifier from the power supply system as a whole. In particular, the magnitude of the PNP leakage current could not be measured. With the low carrier frequencies used in inductive links it is however reasonable to apply a quasi-static first approximation, and assume that the leakage current can be extrapolated from the static behavior. Because of the difficulty in isolating the behavior of the bridge rectifier, the dynamic measurement results have all been put in section 4.4, which treats the performance of the power supply path as a whole.

4.3 Internal linear regulators

Low drop-out regulators have in recent years become increasingly popular, due mostly to the increased efficiency associated with the small drop-out voltage, but also to the reduced headroom needed by such regulators [86, 113]. The reduced headroom allows the supplied circuit to start operating sooner after the power supply is turned on (as in a remotely powered device like a batteryless implant), and allows it to keep on working longer after the power is turned off, or when the battery approaches the end of its discharge cycle.

The current efficiency of regulators is as a matter of course one of the important figures of merit for a low power design. The current into a linear regulator is always equal to or larger than the output current, with the relation between the two depending on the type of regulator. An approximation which describes a wide range of regulator designs is to split the input current into a constant part used internally within the regulator and a variable part which is proportional to the output current. With the branch currents at the regulator terminals defined as in figure 4.31, the relation between the input and output currents is

$$I_{DDH} = I_{int} + (1+\alpha) I_{DDL}, \tag{4.62}$$

and the current efficiency η_I is

$$\eta_I = \frac{I_{DDL}}{I_{DDH}} = \frac{1}{(\alpha+1)} - \frac{I_{int}}{(\alpha+1) I_{int} + (\alpha+1)^2 I_{DDL}} \tag{4.63}$$

The efficiency rises asymptotically to the value $1/(1+\alpha)$ as the output current increases. The fixed current I_{int} should be as small as possible compared to the normal output current to place the operating point close to the asymptote.

Besides the obvious purpose of setting the supply voltage of a subcircuit, a linear voltage regulator can be used to isolate the supply transients originating in the subcircuit from the rest of the system. The typical case where this might be desirable is when the subcircuit consists of a digital logic block which must share a power supply connection with analog circuits. Separate external supplies for digital and analog circuits, with separate pads and routes, are a luxury which is not always available, especially in magnetically powered implants. The supply noise problem can be reduced by adding off-chip filter components and separate power pads, but this introduce undesirable interconnect complexity and bulk. We have therefore investigated a possible way to isolate the digital supply internally on the chip.

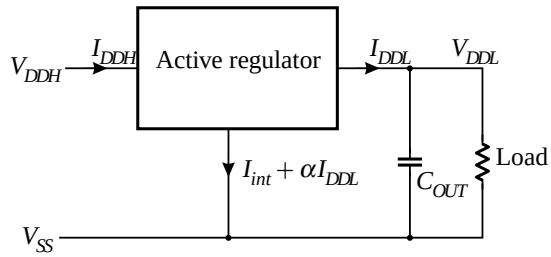


Figure 4.31: *A simple model of a linear voltage regulator. The output impedance at intermediate frequencies is set by a capacitor across the output, which is nearly always present to improve the transient response.*

When a regulator is used as a supply isolation element instead of only as a supply regulating device, the reverse transfer characteristic from the output to the input suddenly becomes important. This transfer characteristic can be described more exactly as the transfer function

$$T(s) = \frac{I_{DDH}(s)}{I_L(s)} \tag{4.64}$$

This is actually the inverse of the definition for the efficiency of the regulator in (4.63). In the context of regulator efficiency, this ratio is however mostly relevant at DC, while the reverse isolation is interesting at higher frequencies. Specifically in the case of noisy digital logic loads, the function $T(s)$ should be minimized at the switching frequencies of the logic and at their harmonics. If the regulator does not contain any energy storage element, this expression will be close to 1 up to relatively high frequencies, and the isolation effect will be nil. It is usually desirable to improve the transient behavior at the output of the regulator by adding a capacitor across the output terminals, and perhaps some internal capacitors in the regulator itself. The impedance of the regulated supply will then have a dominant pole as described before, with the pole frequency determined by the DC output resistance of the regulator and by the capacitor.

4.3.1 Applications of linear regulators

The applications of linear step-down regulators on an implantable IC, or in any system with only one single power supply connection are numerous. The reduction of the supply voltage for analog circuits is rarely a goal in itself since most operate better, faster and with a larger SNR at higher supply voltages. Reducing the power consumption through a reduction in supply voltage is rarely a reason either, as most analog circuits look like current source loads seen from the supply (with the exception of resistive dividers and such, but they can in most cases be scaled with the supply voltage). A valid reason for applying step-down regulators may be to protect analog subcircuits against supply voltages exceeding the technology specification. This requires the availability of a different type of device for the regulator itself, as might be the case in a standard CMOS technology with high-voltage extensions. Some designs for a stimulator chip were in fact carried out in a 0.7 µm CMOS technology with 100 V devices, but they were not fabricated due to technology support issues.

A more compelling reason for inserting a regulator between the supply and an analog subcircuit may be to increase the effective PSRR of the circuit. Power supply transients due to a varying power transmission into the system or due to other circuits on the same chip are especially relevant in the design of implanted systems.

The reverse case is also true, in that a regulator can provide protection for the system supply against transients originating in the subcircuit itself. This reverse isolation feature was one of the main reason for including supply regulators for the logic blocks in our stimulator designs. As we show below, a regulator also decreases the power consumption of a CMOS logic block. This is however not a valid reason to include a stepdown regulator in a system with only a small amount of logic, as the practical difference between current consumptions of 5 and 10 µA may be negligible.

The supply current of some types of logic, among them standard CMOS logic, is strongly dependent on the supply voltage. For standard CMOS logic, where the dynamic current consumption is dominated by the current necessary to charge up node capacitances, the supply current is proportional to the square of the supply voltage [118]. A reduction in the supply voltage can then save power to an extent which depends on the type of step-down regulator. The best power savings are obtained with switching regulator types, but with more practical integrated linear types the reduction in power is proportional to the step-down voltage ratio V_{DDL}/V_{DDH}. The total power reduction is somewhat smaller due to the regulator power overhead, but with a good design it is still possible to achieve a significant reduction in total power consumption. The lower logic supply voltage has the side-effect of increasing the propagation times of logic gates, but with the low clock frequencies (1–5 MHz) used in our designs and with modern technology, the propagation time is irrelevant.

Using the model (4.62) for the internal current use of a linear regulator, we can calculate how large the total supply current of a subsystem consisting of a regulator and a block of CMOS logic is compared to the same logic block without the step-down regulator. Because the load current is proportional to the supply voltage (and the supplied power proportional to the square of the supply voltage), we can define an equivalent load admittance Y_{eq}. With the clock frequency and all other factors being equal, the ratio between the two currents is $I_L(V_{DDL})/I_L(V_{DDH}) = V_{DDL}/V_{DDH}$ In the first case, with no regulator between the supply and the circuit, the current is simply $I_{DDH} = V_{DDH}Y_{eq}$. With the regulator inserted, the supply current is

$$I'_{DDH} = I_{int} + (\alpha + 1)\frac{V_{DDL}}{V_{DDH}}I_{DDH} \tag{4.65}$$

The break-even condition can be found by setting the right side of this equation equal to I_{DDH}, leading to

$$I_{DDH1} = \frac{V_{DDH}}{(\alpha + 1)V_{DDL}}I_{int} \tag{4.66}$$

If the supply current without the regulator is smaller than I_{DDH1}, the addition of a regulator will not reduce the supply current (but one may still be used to reduce the supply current transients). The supply current reduction ratio is shown in figure 4.32 as a contour plot with the normalized internal current consumption and the step-down ratio as coordinates. Given a base supply current for the circuit running directly off the high supply voltage, the graph shows whether using a regulator with a current consumption I_{int} and a given step-down ratio will result in a net reduction in the supply current.

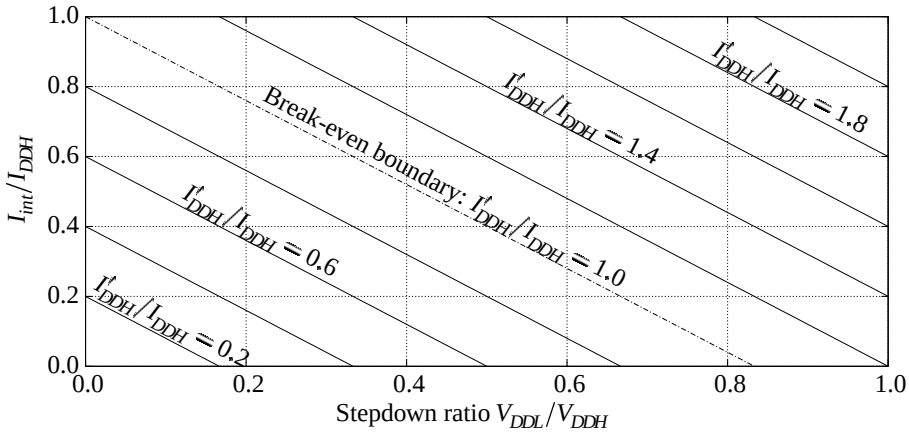


Figure 4.32: *The ratio of the total supply current with a linear step-down regulator to the current without a regulator, when the load is a standard CMOS logic block. The regulator is modeled by the simple expression* (4.62)*. The proportionality constant α is 0.2, corresponding to the regulator implemented in the test circuit.*

4.3.2 Regulator architectures

The regulator presented here was designed for a nominal output voltage of 1.8–2.0 V, and was included in a system where other components required 3.0 V to function correctly. The acceptable dropout voltage was therefore close to 1 V, which is not typical for low dropout (LDO) types. Even so, LDO regulators are frequently aimed at integration in low-power systems, so many of the design goals are compatible with the requirements of implanted systems. It is therefore instructive to compare a few LDO (and not-so-LDO) regulator types with respect to their inclusion in an implanted IC.

A voltage regulator whose primary specification is a low dropout voltage almost invariably consists of a pass transistor between the supply and the load, with the drain or collector towards the load [113, 86], and a feedback loop which regulates the output voltage. This allows operation with dropout voltages in the neighborhood of 150–200 mV if the pass device is configured as a current source. The pass device may, in case it is an MOS, be operated in the linear range, where it is basically a voltage-controlled resistor. This extends the range of possible dropout voltages almost down to zero, with reported values as low as 50 mV [91].

The general design, and specifically the compensation of the feedback loop, will be different in the two cases. Circuits using a current source output need a compensation circuit because the dominant pole in the feedback amplifier is close to the second pole due to the common-source amplifier at the output [97]. The problem is compounded by the fact that is it desirable to use V_{SS} as the common rail, which dictates the use of a p-channel MOS as the pass device, leading to an increase in the load capacitance seen by the amplifier compared to an n-channel device having the same g_m.

To simplify the design and compensation of the regulator one may turn the pass transistor around, and use an n-channel MOS or NPN as a voltage follower. With a follower pass transistor, the regulator can provide good line and load regulation with a low output

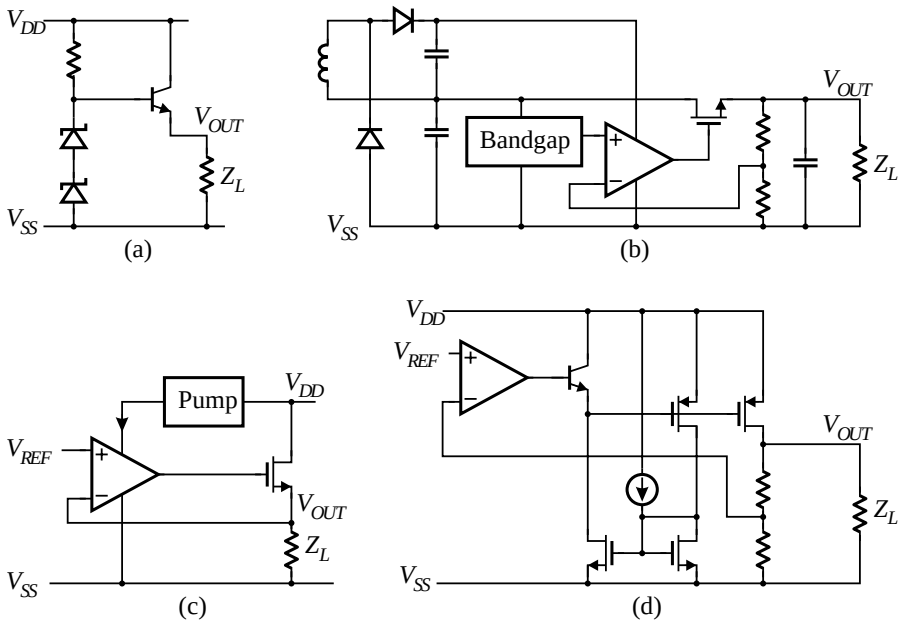


Figure 4.33: *An example of some different linear regulator types. (a) is a simple zener and NPN-based regulator used in an implantable stimulator [122], (b) is a regulator proposed in [91] for use in inductively powered systems, (c) is an NMOS-based voltage follower type with a charge pump to boost the gate drive circuit voltage [97] and (d) is an LDO design with a common source output transistor [86].*

impedance, without requiring any additional compensation since the pole frequency of the follower is significantly higher. The drawback is that the output must be one gate-source (or base-emitter) drop below the output of the feedback amplifier, and the regulator can then hardly be classified as an LDO type (unless the process offers MOS transistors with zero threshold voltage, an unlikely scenario in today's digital-driven technology development). There is however a possible remedy to that problem. Regulator implementations have been reported where the feedback loop itself is supplied by a separate boosted supply which is higher than the system supply [97]. The use of a charge pump to supply the regulator core increases the power overhead of the regulator, but may lead to a net gain, depending on the cost of achieving the same performance by other means.

The regulators described in the literature on implanted electronics have varying degrees of sophistication, with the low-end sector represented by simple voltage followers with zener voltage references. Figure 4.33(a) shows one such example from [122]. While this circuit may perform well enough, and has the virtue of being simple, it requires a zener diode which is available in few processes. Zener breakdown effects are available in some standard CMOS processes, but since they are mostly an uncontrolled by-product their reverse breakdown voltages show large variations. The DC output resistance of the regulator is approximately $R_{out} = R_B/\beta + 1/g_m$ where R_B is the impedance of the base network. Ideally, the output resistance should be dominated by the $1/g_m$ term so that R_{out} can track varying

load conditions. The small-signal resistance of the zener device is therefore important for the performance of the regulator, and the device must accordingly be biased with a high current. The base network shown in the figure was biased with 800 µA, placing it on the wrong side of any subjective boundary of low-power electronics. The zener could be replaced by an active circuit with a sharp bend in the I/V characteristic, but the current is better spent on a different regulator architecture.

A traditional pass transistor design where the pass device is configured as a voltage follower is shown in figure 4.33(b). This circuit was also designed for use in an implant, and it features an opamp in the feedback loop and an internal voltage reference. The circuits in the feedback loop have a higher supply voltage than the high side of the pass device so it is possible to reach a very low dropout voltage. This dropout voltage can be made lower than the saturation voltage of the pass transistor by biasing the gate highly enough, if the feedback network can cope with the altered stability conditions. The higher supply voltage for the feedback loop is created by adding a center tap to the inductive link antenna, and using a voltage doubler rectifier network. It is perhaps questionable whether the use of a voltage doubler in this way generally gives better efficiency than using a simple rectifier network and running the regulator and the load off a single supply.

Another solution which does run off a single supply and benefits from the advantage of increased pass device gate drive is shown in figure 4.33(c). This circuit uses an internal charge pump to create a higher supply voltage for the regulator. In the implementation shown in the reference for this circuit [97], the charge pump and bias circuit used a little over 2 mA, compared to the rated output current of 75 mA. Assuming that the regulator is run with a small dropout voltage so that the full benefit of the high gate drive is obtained, and that the average output current is close to the maximum of 75 mA (a heavy digital load for example), the overhead is quite acceptable.

As a last example, figure 4.33(d) shows a more classical LDO design, with a p-channel MOS pass device which operates as a current source. The increased capacitive load of the PMOS pass transistor compared to the NMOS solution is compensated for by adding a bipolar buffer between the amplifier and the load. The bias circuit for the buffer includes a booster feature which mirrors part of the output current back to the buffer bias, to push the pole associated with the pass device gate out to higher frequencies. This eliminates some of the phase margin penalty due to the increased load of the PMOS. Additionally, the current efficiency is not degraded noticeably because the buffer bias current is a small fraction of the total current under all operating conditions. A necessary condition for the operation of this circuit is that the PMOS threshold voltage may not be smaller than the NPN's V_{BE}. In a deep submicron CMOS process with threshold voltages on the order of 0.5 V, or if the operation amplifier output can not reach close enough to the positive rail, there may not be enough headroom for a bipolar buffer. Replacing it with an NMOS may be an option depending on the relative gate-source voltages of the two transistors, taking the large bulk modulation of the NMOS transistor into account.

The regulators discussed up to now have some attractive features, but they represent the wrong kind of tradeoff for the specific task at hand: they provide low dropout voltages or low output resistance at the cost of increased current consumption, or they require advanced process features. To supply the local logic blocks in an implanted IC the requirements are a bit different. The system-wide supply voltage is not set by a battery but by the technology limit, typically 2–3 V in a submicron CMOS process. As the supply current of the logic load

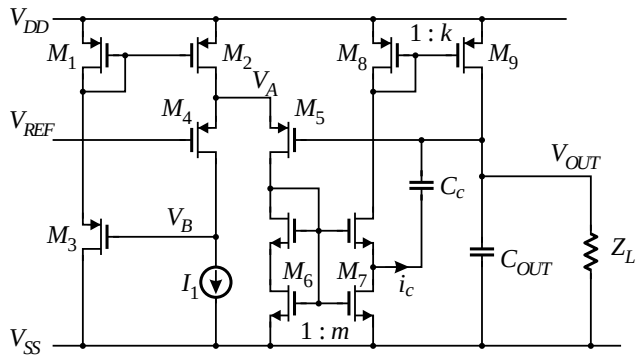


Figure 4.34: *A low-power regulator from [95] which provides a very wide output current range.*

drops with falling supply voltage and the logic will function down to supply voltages below 1 V, a low dropout voltage is not an important consideration. Dropout voltages on the order of 0.5–1.0 V can easily be accommodated without compromising the total power budget or the logic functions. The DC output resistance is likewise irrelevant since the digital supply is stressed mostly at high frequencies where the output resistance is determined by the output capacitor.

4.3.3 A regulator proposal

Figure 4.34 shows a low-power regulator from [95] which is almost ideal for the task at hand. It was originally proposed as a regulator capable of providing a very wide output current range (5 decades) with minimal quiescent current. The regulator consists of a level shifting buffer M1–M4 and a second buffer M5–M9. The transistors M4 and M5 may be viewed as a differential pair, or more instructively, as two level shifters in cascade. The feedback loop in the second buffer maintains the output at one gate-source potential below V_{s4}, or approximately equal to the reference voltage. By splitting the transistors in the first current mirror M6–M7 of the second buffer to provide a tap in the middle of the channel, it is possible to use the fact that the small-signal impedance is lower towards the "top half" of the transistor, and provide a current path for compensation of output transients. In case of a negative voltage transient on the output the current through C_c and M7 is amplified and mirrored back to the output.

These qualitative observations give a general idea of the operation of the regulator and the features which improve its transient response. A quantitative analysis is however necessary to understand how the dynamic response of the regulator is controlled. As the regulator contains two connected feedback loops the analysis is not quite trivial. This is as good an excuse as any to show a nice small-signal diagram, so figure 4.35 shows one of those. The diagram is derived directly from figure 4.34 and needs no further explanation except for the way in which the compensation capacitor C_c and the two parts of M7 interact. The elements C_A, C_B and g_B represent parasitic capacitances and conductances at the nodes marked by the voltages V_A and V_B. The notation is simplified by denoting the total small-signal input capacitances of the current mirrors by C_{g1}, C_{g6} and C_{g8}.

The top segment of M7 is in saturation (assuming that the total transistor is saturated) and

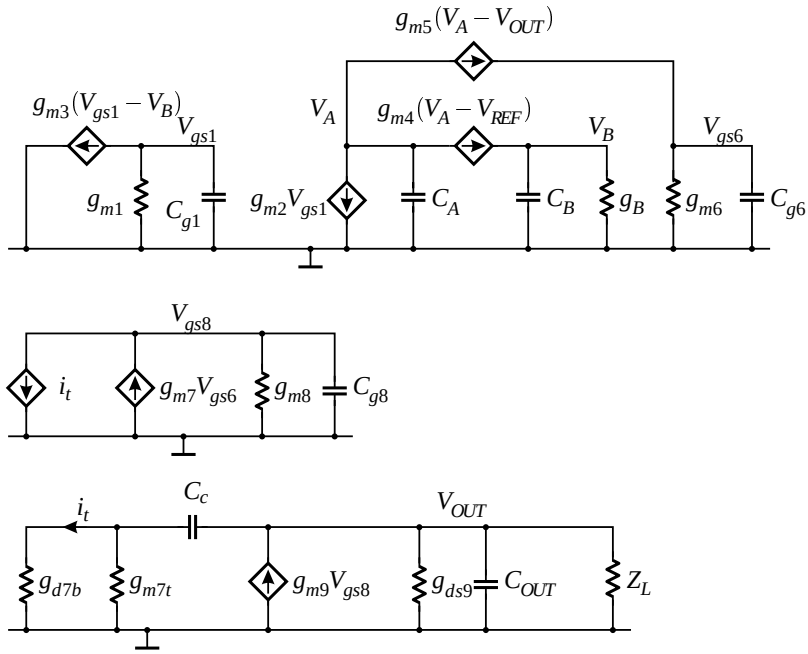


Figure 4.35: *A small-signal diagram for the regulator in figure 4.34. The capacitances C_{g1}, C_{g6} and C_{g8} denote the total capacitances between the current mirror inputs and small-signal ground.*

the bottom segment is in the linear region. The small-signal conductance looking into drain of the bottom (M7b) segment is then $g_{d7b} = \sqrt{2\beta_b I_{d7}}$ and the resistance into the top (M7t) segment is $g_{m7t} = \sqrt{2\beta_t I_{d7}}$. These expressions are based on the square-law Schichmann-Hodges model for lack of another more accurate analytical model. The compensation current i_c through C_c will then be split according to the aspect ratios $A = W/L$, with a ratio

$$\frac{i_t}{i_c} = \frac{\sqrt{A_t}}{\sqrt{A_t} + \sqrt{A_b}} \tag{4.67}$$

passing through the top transistor. To get the maximum benefit from the compensation mechanism the aspect ratio of the top transistor should be large compared to the bottom transistor. Assuming that the conductance of the top segment dominates, the 3-dB frequency of the compensation path is

$$\omega_c = \frac{g_{d7b} + g_{m7t}}{C_c} \tag{4.68}$$

We can observe in passing that the top segments of M6 and M7 could just as well have been biased separately as low-voltage cascodes. The fraction i_t/i_c would then have been close to 1 and the output impedance of the mirror improved, at some extra cost in a bias circuit.

The full transfer function from the output current to the output voltage and from the input reference to the output are a bit unwieldy because they contain 5-6 poles. Some of those can be neglected for a first round of analysis, and the interaction between the two stages can be

eliminated by cutting the signal path at the source of M5. Manipulation of the small-signal equations for the first buffer leads to the results

$$H_1(s) = \frac{v_A(s)}{v_{REF}(s)} = \frac{g_{m2}g_{m3}}{(g_{m2}g_{m3} + g_B(g_{m1} + g_{m3})) + s(g_{m1} + g_{m3})C_B} \tag{4.69}$$

and the output conductance is simply g_{m4} boosted by the open loop gain,

$$G_1(s) = -\frac{i_{s5}(s)}{v_A(s)} = -g_{m4}\left(1 + \frac{g_{m2}g_{m3}}{(g_{m1} + g_{m3})(g_B + sC_B)}\right) \tag{4.70}$$

These were obtained by setting all capacitances to zero, except for C_B which determines the dominant pole frequency of the buffer. For a stability analysis, the node capacitances at the input and output of the current mirror M1–M2 must also be taken into account. Their effect can be controlled by increasing C_B so that the mirror pole frequencies are a fair margin above the unity-gain frequency of the open-loop gain of the buffer.

The response of the second, current amplifier-based buffer, can be estimated at low frequencies, again by ignoring the mirror poles and disconnecting the two buffers. If the load impedance Z_L is replaced by a current load I_L, we can easily calculate the output resistance of the regulator. The open-loop DC impedance level at the output of the regulator is $R_o \approx r_{ds9}$ since the output resistance of the (small) bias current source is much larger than that of M9. Neglecting internal poles, the output voltage is then

$$v_{OUT} = \frac{1}{1 + s/\omega_0}\left(v_A - \frac{1}{mkg_{m5}}i_L\right) \tag{4.71}$$

where the pole frequency

$$\omega_0 = \frac{R_o mkg_{m5} + 1}{R_o C_{OUT}} \tag{4.72}$$

is moved out by a factor mk from the open-loop case, as might be expected. To obtain the output resistance of the entire buffer we may replace v_A in (4.71) by an expression which includes the output resistance of the buffer as in (4.70), but the change in the value of the output impedance is small, and taking into account the other approximations used in the derivation, it is not worth the effort.

The transfer function of the full regulator circuit is rather complex, and no particular purpose is served by deriving it here. Instead we observe that at the DC bias levels expected from a digital load which does not go into a power-down mode, the compensation capacitor C_c is not necessary. The purpose of the capacitor is to improve the transient response when the output current increases suddenly by orders of magnitude, and to reduce the high-frequency small-signal output impedance. The reduced output resistance obtained by the inclusion of this capacitor is not desired in our case since it would counteract the isolation of the logic load from the system supply.

There is one problem with the regulator circuit which was not identified in the original source. The regulator will only function correctly over a limited input voltage range because of the design of the reference voltage buffer. The gate of M3 is connected to the positive supply through two gate-source connections and is therefore maintained at a relatively constant voltage below the input. As the input voltage increases the gate potential of M3 increases relative to V_{REF} until M4 goes out of saturation and the feedback loop is broken.

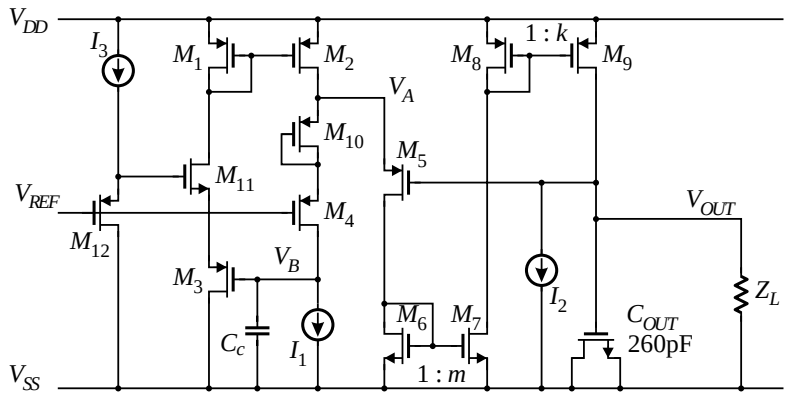


Figure 4.36: *A modification of the regulator in figure 4.34 which does not suffer from the limitation in the input voltage range.*

The problem can be remedied by adding a transistor between M3 and M1, to transform the impedance driving the gate of M1 from low to high. The additional transistor, which is marked as M11 in figure 4.36, functions as a basic CMOS current conveyor [19]. Without the impedance shift, the regulator will only function with input voltages up to $3|V_T| + 2|V_{DS,sat}|$ above the reference input. Since the design also imposes a lower limit on the input voltage, determined by V_{gs4} and V_{ds2}, the input voltage range of the original regulator only spans two threshold voltages. As the simulation in figure 4.37 shows, the original circuit will not function correctly up to the top of the normal operating range (3.3 V in this case) but the modified circuit has no such limitation.

The large output capacitor in the regulator in figure 4.36 was as the figure shows, implemented using a MOS gate. The nonlinearity inherent in the gate capacitance is not important since its only effect is to vary the transient response slightly according to the DC bias level. As the gate-bulk voltage is about 2 V, and the threshold voltage is 0.5–0.6 V, this effect is undetectable. Another possible concern besides the nonlinearity is that the high-frequency response of the capacitor might be affected by the series resistance in the gate and the bulk. The MOS was therefore laid out in a fingered structure to reduce the resistances to a level where the capacitor's own time constant is above 100 MHz.

To obtain an output voltage of approximately 2 V with the global reference voltage of 1.25 V a level shifter M10 was inserted at the source of M4. A minimum load current of $I_2 = 500$ nA was also added to provide a minimum bias for the current amplifier. This current load replaces the function of the original compensation capacitor C_c without altering the high-frequency response of the regulator.

With the implementation shown in the diagram, the regulator can only provide output voltages up to 0.6 V below the input. As was explained above this is not a real limitation since a low dropout is not required. If it were, it could be implemented by adding a level shifter between the output and the gate of M5.

Table 4.1 shows a comparison between the regulator designs presented in this section. The table is only intended as a rough overview, since it is not necessarily meaningful to compare regulators implemented in different technologies and with different purposes. The

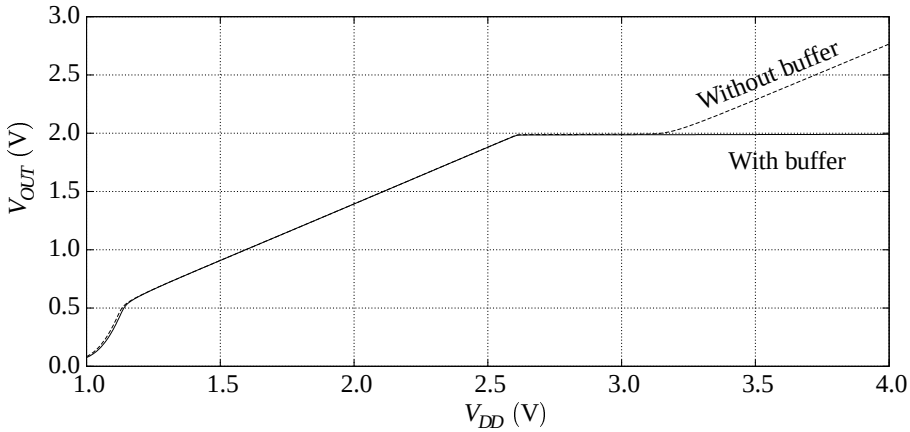


Figure 4.37: *The simulated output voltages of the regulators in figures 4.34 and 4.36. The effect of the additional transistor M11 in the latter is visible in that it removes the upper limit on the operating input voltage range.*

Ref.	Technology	V_{DD} (V)	V_{OUT} (V)	I_{int} (µA)	α	V_{DO} (mV)	I_{load} (μA)
[97]	0.25 µm	3.3	2.5	2300	0.00	200	75000
[113]	Bipolar	1.1–1.6	0.9	46	-0.12	200	140
[91]	0.8 µm BiCMOS	5.0	4.0	240	0.00	245	5000
[86]	2 µm BiCMOS	1.2	0.9	23	0.004	100	50000
[50]	2 µm BiCMOS	12	3.0	50	0.00	200	200
[46]	2.4 µm	12	3.3	54	0.00	1200	1000
Figure 4.36	0.5 µm	3.0	2.0	1.5	0.20	600	100

Table 4.1: *A comparison of published linear regulators and the regulator presented here. The load currents in the last column are the maximum figures.*

internal supply current has for example been traded off in some cases for increased bandwidth, while we have decided to let the output capacitor control the output impedance down to relatively low frequencies.

The factor α in (4.62), which describes the proportional part of the current consumption, depends in this regulator mostly on the ratio $1 : k$ of the last current mirror. The mirror ratio depends in turn on the stability criteria for the entire regulator, with larger mirror ratios being harder to compensate. In the present case, the current overhead at the nominal load was only 1–2 µA with $k = 6$.

4.3.4 Measurement results

The test chip with the regulator on it incorporated a few test access pads for basic functionality measurements. A regulator of this type does not usually need to drive off-chip loads and no special features were added to it to improve the output characteristics at the test points. The accessible terminals were the reference voltage input and the output, through ESD protection circuits. The supplies were shared with other circuits on the IC.

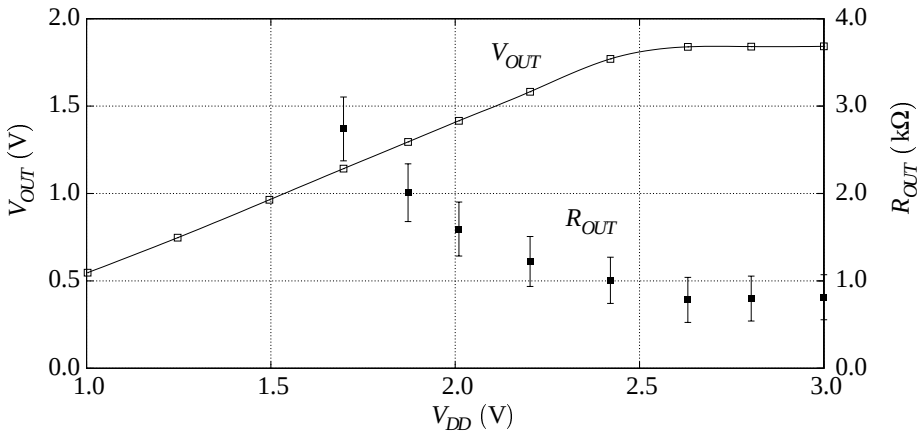


Figure 4.38: *The measured output voltage of the regulator in figure 4.36 and its output resistance as a function of the input voltage, with a 100 kΩ load.*

The response of the regulator to line and load variations was only investigated at DC, both because of limitations of the test setup and because of the lowpass effect of the ESD circuits. The 3dB frequency introduced by the protection circuit was for example about 2-3 times lower than the (presumed) 3dB frequency of the output impedance of the regulator.

This kind of limitation in the access to information regarding the internal functioning of the IC is really due to the strategy of integrating many interlocked subcircuit on a test chip to form a complete system. Instead, one could have chosen to place each block separately with full external access. That would have made it impossible to test the system as a whole because of the effect of inserting off-chip paths into the signal path. With relatively straightforward circuits like this regulator, simple functionality tests and DC measurements provide enough information to determine with a high level of confidence, whether the circuit is working according to specifications.

The dropout voltage of the regulator is shown in figure 4.38, with a 100 kΩ load (corresponding to over twice the nominal output current at the target output voltage of 2 V). The output voltage is seen to be slightly below the target value but the difference can be traced to process variations and transistor mismatch (which are the same thing at some level). The dropout voltage of about 0.6 V is due mainly to the gate-source voltage of M5. The falling output resistance with increasing output voltage is a result of the increasing load current, which in turn increases the transconductance of the buffer transistors.

4.3.5 Isolation performance

One aspect of the isolation performance of the regulator is governed by the paths through which the output node is connected to the supply and how the load current affects the current drawn from the supply. The drain current of M9 is obviously the largest contributor to the supply current, at least at DC. The output capacitance of the last mirror was absorbed into the output capacitor in the small-signal diagram in figure 4.35 since we were not concerned with the origin of the output current. In contrast, for the isolation performance calculation,

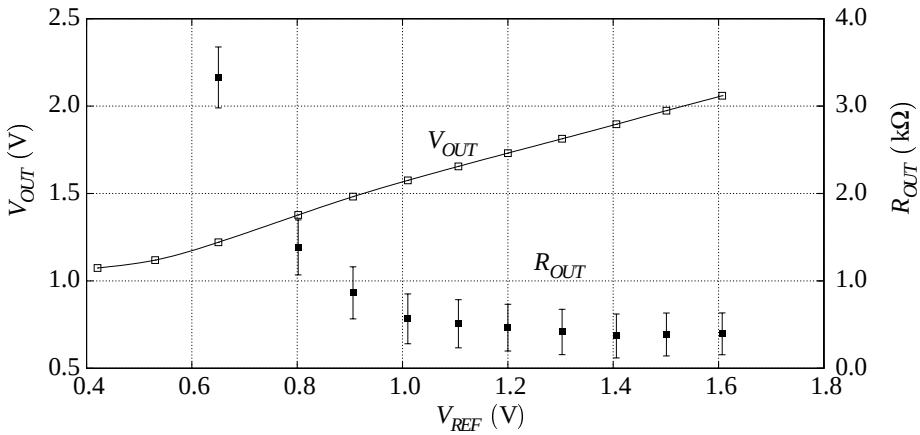


Figure 4.39: *The measured dependence of the regulator output on the reference voltage. The input voltage was $V_{in} = 3\,V$. At the nominal reference voltage of 1.25 V the regulator output is 1.8 V, slightly less than the target value.*

the drain-bulk capacitor of M9 is critical since it is the main feedthrough path at high frequencies.

The isolation is characterized by the transfer function $T(s) = I_{DD}(s)/I_L(s)$, with the supply current I_{DD} consisting mainly of i_{d8}, i_{d9} and v_{OUT}/sC_{db9}. This function can be calculated but the resulting expression does not lead to much insight. Instead we can observe that the current $i_{d9} + i_{d8}$ is equal to $(k+1)/k$ times the output current up to the pole frequency ω_0 given in (4.72) where it starts falling by 20dB per decade. At high frequencies, where the output impedance is mostly capacitive, the output current is split between C_{db9} and C_{OUT}. The high-frequency value of $T(s)$ is therefore $C_{db9}/(C_{db9} + C_{OUT})$, at least until other parasitic phenomena like the output capacitor's internal resistance begin to dominate.

From the previous observation we see that the isolation performance of the regulator can be improved in several ways. The corner frequency of $T(s)$ can be lowered by increasing C_{OUT} as much as the available die space allows. Decreasing the transconductance of M5 and the mirror ratios m and k will also lower the corner frequency, but it will also increase the DC output resistance, so a compromise must be found. The high-frequency value of $T(s)$ can be reduced by making the drain-bulk capacitance of M9 as small as possible, for example by using a narrower transistor at the cost of some more gate-source voltage. Alternatively, it is possible to add a cascode transistor to M9 with a floating well to reduce the parasitic capacitance from the output to V_{DD}.

An equivalent low pass filter section can be defined as in figure 4.41 to approximate the output resistance and the supply loading characteristics of the regulator. This simple model does not account for the effect of the internal poles and zeros but it can be used as a design aid. The full transfer function $T(s)$ must still be evaluated using a circuit simulator as in figure 4.40.

The supply noise from a digital load is mainly concentrated at the clock frequency and at its harmonics so the important performance figure for the simulator is the isolation at that frequency. The clock frequency was 1.6 MHz for the system in which the reported

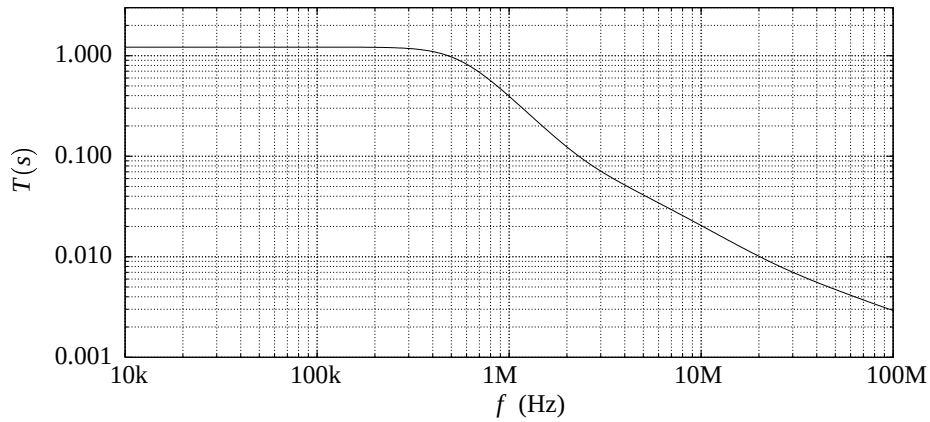


Figure 4.40: *The simulated supply output/input isolation performance $T(s)$ of the regulator in figure figure 4.36.*

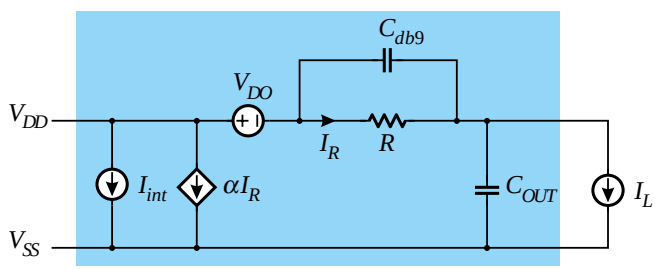


Figure 4.41: *An equivalent low-pass section for the regulator which approximates the output resistance and the supply loading of the regulator. The two current sources on the supply side account for the current overhead of the regulator.*

regulator was integrated and the simulation results show that the isolation at that frequency is approximately 8dB. This rather dismal figure can be improved by increasing the output capacitor, or by decreasing the product mkg_{m5} at a cost in DC output resistance.

Another possibility is to use more than one regulator of this type in cascade. The supply headroom and extra supply current overhead may however make this method impractical. There are also other paths through which the supply can be contaminated, which have not been included in this simple model. If the direct path through the regulator is reduced by a large margin, other noise mechanisms like substrate coupling are likely to dominate. This design was implemented in standard CMOS with a highly conductive substrate, providing the ideal conditions for substrate noise coupling. The effect of the coupling can be reduced providing for each analog block a local decoupling capacitors which is much larger than the parasitic V_{DD} to V_{SS} capacitance [35]. Since this capacitance consists mostly of n-well to bulk junction capacitances and to a lesser degree of V_{DD} routing capacitances, the amount of decoupling can be impractically large on a mostly analog chip.

4.4 Supply quality

The quality of the power supply seen by circuit blocks in an implanted IC is influenced by the entire power transfer path from the external transmitter through the link and onto the chip. The parts along the way which contribute to the characteristics of the supply voltage seen by a single circuit block are the transmitter architecture, the modulation method, the transmitter and receiver antennas and the coupling between them, the power converter, the energy storage capacitor, the overvoltage protection and supply transients from other subcircuits. We have in the preceding sections examined the most important of these elements, and shall attempt to provide here a picture of the complete supply quality.

4.4.1 Contributing factors

The blocks which are most sensitive to supply noise are analog signal processing blocks. The supply noise seen by such a block depends on the power input path into the energy storage capacitor and the load transients generated by other loads. In that respect, the high-frequency digital and analog loads are the worst offenders. The impedance level of the on-chip supply is low at DC due to the action of the shunt regulator. We shall assume below that the overvoltage limiting circuit is active, since there is only a narrow operating region between the minimum supply voltage set by the design and the technology limit (at least in the 0.5 µm CMOS design that most of these results are based on). The strong dependence of the transmitted power on the link distance means that in practice, the transmitted power is larger than required.

The power converter, or in other words the bridge rectifier, is the largest noise contributor. Since the supply noise from the rectifier is concentrated at the harmonics of the carrier frequency it should be possible to filter it out in signal processing stages or eliminate it with differential topologies. Because of its large amplitude compared to some of the signals this can however be problematic. Slight mismatches in differential branches or intermodulation by nonlinearities in the signal path may lead to problems if they are not controlled carefully.

The digital logic circuits in the system are another large noise contributor, and since the system clock is often derived from the carrier frequency, the logic noise is also concentrated at (fractional) harmonics of the carrier frequency. The logic noise spectrum is more complex than that of the power converter since the spectrum is modulated by changing internal states and activity levels.

Figure 4.42 shows simulation results for the supply current of the control logic block on a transceiver chip (see section 6.2). The simulation was run over a long period of time so that the logic block moved between several activity levels during the simulation interval. Despite varying activity levels, the spectrum shown in the figure differs very little from what one would expect from an unmodulated pulse train at twice the clock frequency. The frequency components corresponding to the clock frequency itself are suppressed, mainly because of the near equality of the current transients on both clock edges.

4.4.2 Supply path functionality

In order to test the complete supply path from the transmitter to the internal supply voltage a transmitter and receiver were set up using the power supply conditioning circuits described

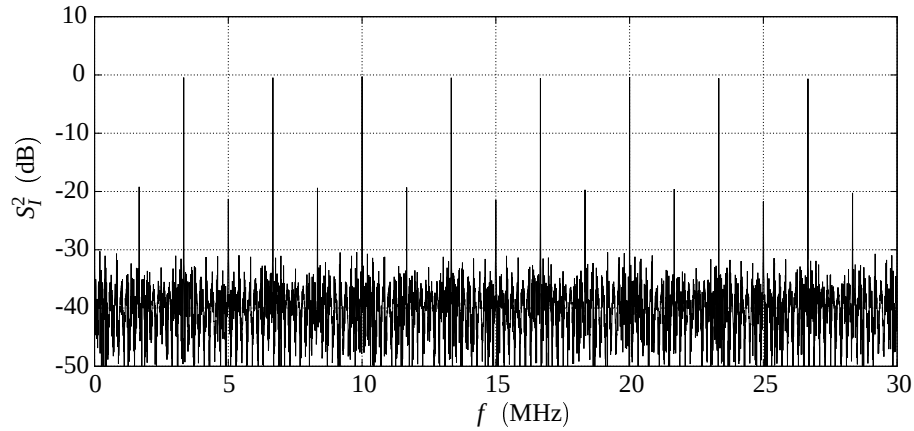


Figure 4.42: *The supply current power spectral density for a complete logic block including state machine, counters and data registers. The system clock period was 1.666 MHz. The spectrum is normalized to the largest component, which is at DC. The time-averaged supply current is 6.8 µA.*

in this chapter. The transmitter was a class D switching type driving a series resonant load consisting of a 430 pF high-voltage capacitor and a 5-turn, 56 mm diameter coil with an inductance of approximately 2.6 µH. The transmitter carrier frequency was 4.64 MHz. Due to limitations in the power MOS gate drive circuit in the transmitter, it was not possible to run the transmitter at higher frequencies (at 6.78 MHz for example for operation in the ISM band) without overheating.

The secondary circuit is a parallel LC circuit consisting of a 470pF capacitor and a 6-turn, 37 mm diameter coil made of 0.4 mm diameter copper wire. The losses in this inductor are smaller than that which can be expected from implanted coils, which have a smaller diameter and more turns of finer wire, and possibly another conductor material. In order to imitate realistic conditions, the Q of the resonant circuit was lowered to 3.1 at 4.6 MHz by adding an 1800 Ω resistor in parallel. This choice of Q is perhaps too conservative since almost any coil implementation will have a better Q, but it does provide a check of the behavior under worst-case conditions.

The system under test consisted of the bridge rectifier, shunt regulator and the stepdown regulator described previously in this chapter. The system also included other circuits on the same chip for a total load current of approximately 150–200 µA. The transmitter which was used for the test run was rather overpowered for the task. It is a very powerful transmitter which was built to make measurements with weakly coupled links, and for the purpose of this test it was run at the lowest power level at which it would oscillate. Despite this, the current amplitude in the primary was 1.4A, giving 7.0 Ampère-turns in the transmitter coil. At this power level, the transmitter drew 0.8A from a 6 V supply. The two coils were arranged on a common axis and the on-chip supply voltage was measured as a function of the distance between them. The results in figure 4.43 show that the full operating supply voltage is attained at a distance of 120 mm which is ample for most applications. A more typical distance for a shallow implant would be 20 mm. The excess power available from the

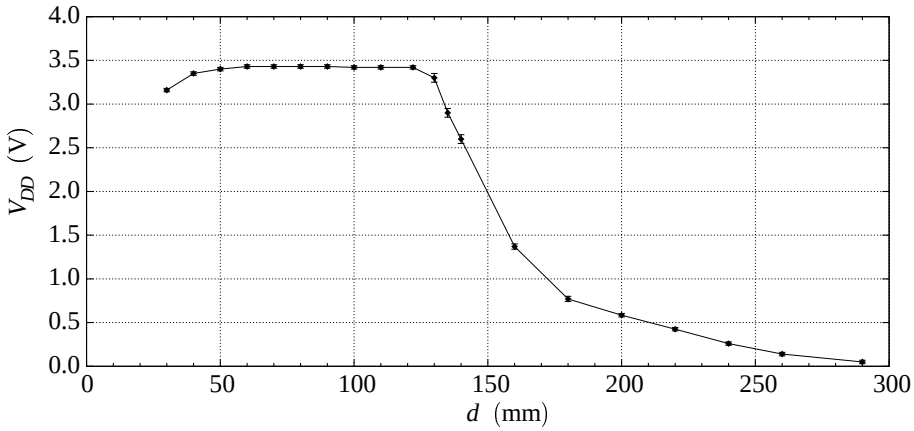


Figure 4.43: *The rectified and regulated on-chip supply voltage results from a test run with an inductive link. The shunt regulator limits the supply voltage to 3.3–3.4 V at all distances.*

link at smaller distances than 120 mm serves to show that the shunt regulator can sink enough current to regulate the supply voltage under extreme conditions. The shunt current could not be measured directly, but the DC measurements in section 4.1 show that it is functional up to 80 mA. Figure 4.43 shows a dip in the supply voltage at very small distances. This dip is due to the fact that the coupling between the primary and secondary becomes high enough to alter the resonant frequency of both circuits, and the transmitter does not compensate for the frequency shift.

It is interesting to compare this result for the maximum range of the link to the analytic expression derived in section 2.2. By applying the data from this test run to (2.27), we obtain a maximum range of 105 mm. This is quite close to the measured result of 120 mm, considering the approximations made in the derivation of (2.27).

5
REFERENCE CIRCUITS

In a self-contained implanted electronic system, with a minimum of components besides the main system IC, problems arise which are not usually encountered in board-level systems. These problems are due do the lack of external system resources like reference voltages and bias currents, time references and supply and ground connections. Many of the external services which can be provided by separate ICs do not exist in our reference case since the system consists only of a single IC and the smallest practical number of passive discrete components. This is true whether the system is a single-chip nerve stimulator IC or a distributed transducer system consisting of two or more interconnected chips. Interconnects are a very expensive resource in implanted systems and all reference signals which can be generated locally, should be. We shall in the following present solutions to some of the reference circuit problems in an implanted system IC. CMOS is again used as the reference technology.

Providing bias signals for on-chip analog blocks is a problem in itself, because of the difficulty in setting a predefined bias current subject to normal processing variations in component values. It is in fact impossible to set an internal bias current which does not vary as much as the process variations without resorting to an external component. A bandgap voltage reference can be used to set internal voltage levels quite accurately and predictably, but constant-voltage biasing of analog blocks is rather useless since the resulting biasing currents are subject to threshold voltage and gain factor variations in addition to quite severe temperature effects. To translate the bandgap voltage into a specified bias current it is necessary to use a resistor, and internal resistors suffer from the usual 10–20% process variations. The popular solution is to use a so-called constant-g_m bias cell [116, 63] which generates a bias current with the same variations as a single on-chip resistor, but is insensitive (to first order) to temperature.

Digital logic on a system chip will start up in an unpredictable state when the power is turned on. A reset mechanism is therefore necessary to place the circuit into a known initial state after start-up. This is handled in most digital systems by an external reset control which monitors the system clock and supply voltage, and generates an appropriate reset signal or sequence of signals depending on the sensed conditions. In the case of a single-chip implant this must be handled by the system itself. Since the system clock is also generated locally there is a potential for feedback situations between the reset circuit and the clock generator.

An important task in the design of the clock and reset generators is therefore to anticipate this kind of phenomenon and if possible, design one of the systems to be completely autonomous (typically the clock generator). The supply thresholds at which the clock and reset generator are activated are also important. In a scheme where the clock generator is independent of the reset signal, but where the reset generator monitors the clock signal, the clock generator should start running at a supply voltage which is some margin below the trigger threshold for the reset signal.

5.1 Bandgap references

The voltage reference is a very useful circuit block which finds a wide range of applications. In implanted circuit design, voltage references can be used to set voltage regulator output levels, receiver detector levels, stimulation pulse parameters, etc. Some of these applications only require a reproducible reference voltage signal instead of a specific voltage level. Since there are typically many analog blocks on a system chip which require a reference voltage it is economical to share a single reference between them. Some of those analog blocks, like the voltage regulators, do benefit from a well-defined reference voltage so a common voltage reference should fulfill that requirement.

The literature contains many examples of voltage references which are based on threshold voltages or forward p-n junction voltage drops. These do however suffer from poor repeatability and temperature variations, so they cannot be used in applications requiring specific reference levels. This is rather unfortunate from one point of view since a forward-biased p-n junction for example give the lowest reference source impedance (U_T/I_B) and noise PSD (the wideband noise being just the shot noise $S_I = 2qI_B$) per unit bias current.

The universal solution to the voltage reference problem is to use a bandgap voltage reference. The band gap[1] is the energy difference between the valence band and the conduction band, and is on the order of 1eV in semiconductor materials. The bandgap energy is a nonlinear function of temperature, and is around 1.17eV in pure silicon at $T = 0\,\mathrm{K}$, falling to 1.12eV at $T = 300\,\mathrm{K}$ [10]. The fact that bandgap reference circuits do not measure the true gap at $T = 0\,\mathrm{K}$, but rather the linear extrapolation from the operating temperature to $T = 0\,\mathrm{K}$, means that the bandgap voltage usually refers to the extrapolated voltage which is $V_{G0} = 1.20$–$1.25\,\mathrm{V}$. It should also be pointed out here that since the voltage V_{G0} is a linearly extrapolated value from a nonlinear function of T, the value depends on the target operating temperature T_0.

5.1.1 Basic structure

All bandgap references are based on the summation of a PTAT (proportional to absolute temperature) voltage and another voltage which has a negative temperature coefficient. The negative temperature coefficient is obtained by using the base-emitter potential of a bipolar transistor biased at a constant current, or the forward drop of a p-n junction. The explicit thermal voltage term U_T in the expression for the base-emitter voltage drop $V_{BE} = U_T \ln(I_C/I_S)$

[1] Following the standard solid-state circuits nomenclature, we contract the words “band gap” into a single word in the following.

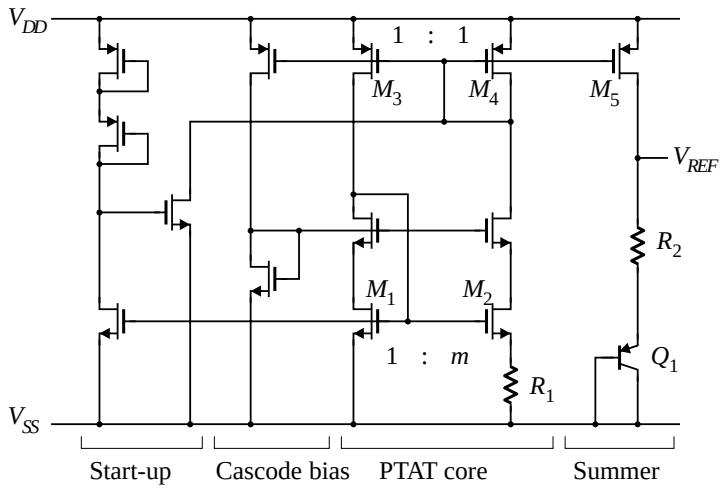


Figure 5.1: *A CMOS-compatible bandgap voltage reference. The transistor Q1 is a bulk PNP where the base is an n-well, the emitter is a p-type diffusion in the well and the p-type bulk is the collector.*

does not account for the entire temperature dependence since the scale current I_S also depends on the temperature [41]. The well-known rule of thumb is that V_{BE} decreases by approximately 2 mV per Kelvin, but the exact rate depends on processing details and device geometry. The dependence of V_{BE} on temperature is described in great detail in [108].

The PTAT voltage source is necessary to compensate the negative temperature coefficient of the V_{BE} voltage. If the temperature coefficient of the PTAT source is set correctly, the sum or average of these two voltages is constant with respect to small temperature variations and the sum of the two is equal to the base-emitter voltage extrapolated to absolute zero. PTAT voltage sources have been implemented in many ways since the bandgap reference was first introduced [119], and we shall not attempt to list them here.

The simple addition of the V_{BE} and PTAT voltages results in a reference voltage which is independent of temperature only to first order. In high-precision references, this is usually not good enough and special circuits are used to eliminate higher-order error terms [85, 11, 48]. This is not necessary in our application, both because the encountered temperature variations are relatively small and because the required precision is not high. Furthermore, high-precision references almost invariably require post-process trimming by laser or by other means to achieve their precision rating [42].

Another development direction for bandgap references is towards lower supply voltages, in some cases lower than the bandgap voltage. There is nothing paradoxal about this since the bandgap voltage does not have to be present anywhere in the circuit for certain reference topologies [120, 101]. Since our design is aimed at systems with a supply voltage of at least 2–3 V, special low-voltage techniques need not be used.

The reference circuit used here is basically the CMOS design presented in [110] with a few modifications. The design is shown in figure 5.1. The PTAT core of the reference is based on a degenerate NMOS current mirror M1–M2, with a feedback path consisting of a PMOS mirror M3–M4. If the NMOS mirror is biased deep in weak inversion, and the PMOS

mirror is ideal, it can be shown [116] that the current in the M2–M4 branch is

$$I_{D2} = \left(\frac{U_T}{R_1}\right)\ln(m) \tag{5.1}$$

This current is PTAT since $U_T = kT/q$. The only process-related physical parameter in this equation is the resistance R_1. The temperature coefficient of the PTAT current is modified somewhat by the temperature coefficient of the resistor, but this effect can be ignored if high-quality resistors are available in the process. In critical work, the temperature coefficient of the resistor can be included in the analysis.

5.1.2 Design considerations

If the PTAT current is to behave according to (5.1), the transistors M1 and M2 must be biased in weak inversion. Since M2 is wider than M1 by a factor $m = W_2/W_1$ and they carry the same current it is M1 which must be dimensioned for weak inversion. There is no single valid boundary for the weak inversion region, and it is not always clear when the transistor dimensions are chosen how much is enough. It depends on the context and the level of acceptable error. Some guidelines do exist, and the criterion used in [110] is that the drain current should be smaller than

$$I_D \leq \left(\frac{n-1}{e^2}\right)\left(\frac{W}{L}\right)\mu_n C_{ox} U_T^2 \tag{5.2}$$

This translates into $I_D \leq 2.5\,\text{nA} \times W/L$ in the 0.5 µm technology used for this reference. This can be compared to the inversion coefficient $ic = I_D/(2n\frac{W}{L}\mu_n C_{ox} U_T^2)$ defined in [34], where the limit for weak inversion is simply set at $ic \ll 1$. As a working tool, this definition is of rather little help. We can however observe that the normalizing current in the definition of the inversion coefficient is larger than the limit in (5.2) by a factor $2ne^2/(n-1)$, which is 50–90 for n in the range 1.4–1.2. We can therefore assume that if we obey the limit in (5.2), the inversion coefficient condition is fulfilled.

Cascodes were added to the NMOS core transistors since they have short gate lengths, and therefore a relatively low output resistance. The p-channel transistors in the top mirror have very long gates (20 µm) and need no cascodes.

The addition of the PTAT voltage and the V_{BE} voltage is carried out in the R2–Q1 branch in figure 5.1. The current mirrored from the PTAT core through M5 into this branch is

$$I_{D5} = \left(\frac{W_5}{W_4}\right)\left(\frac{U_T}{R_1}\right)\ln(m) \tag{5.3}$$

and the output voltage is

$$V_{REF} = I_{D5}R_2 + V_{BE} \tag{5.4}$$

An expression for the base-emitter current which accounts for temperature effects is [18]

$$V_{BE} = V_{G0}\left(1 - \frac{T}{T_0}\right) + \left(\frac{T}{T_0}\right)V_{BE0} + pU_T\ln\left(\frac{T_0}{T}\right) + U_T\ln\left(\frac{I_C}{I_{C0}}\right) \tag{5.5}$$

where p is a fabrication constant for the transistor, and V_{BE0} and I_{C0} define a single operating point at a temperature T_0. The output voltage of the reference is then

$$V_{REF} = U_T \frac{R_2 W_5}{R_1 W_4} \ln(m) + V_{G0}\left(1 - \frac{T}{T_0}\right) + \left(\frac{T}{T_0}\right) V_{BE0} + pU_T \ln\left(\frac{T_0}{T}\right) + U_T \ln\left(\frac{I_C}{I_{C0}}\right) \quad (5.6)$$

The condition for zero temperature coefficient can be found by taking the derivative of this expression and setting it to zero at $T = T_0$. The last term in (5.5), which accounts for the base-emitter voltage dependence on the bias current, was ignored in [110], resulting in the condition

$$\frac{R_2 W_5}{R_1 W_4} \ln(m) = \frac{V_{G0} - V_{BE0}}{U_{T0}} + p \quad (5.7)$$

We can include the last term in (5.5) by using (5.3) for the current I_C, and substituting

$$U_T \ln\left(\frac{I_C}{I_{C0}}\right) = U_T \ln\left(\frac{T}{T_0}\right) \quad (5.8)$$

The condition then becomes

$$\frac{R_2 W_5}{R_1 W_4} \ln(m) = \frac{V_{G0} - V_{BE0}}{U_{T0}} + p - 1 \quad (5.9)$$

By inserting this condition into the definition of the output voltage (5.6) we obtain

$$V_{REF}(T_0) = V_{G0} + (p-1)U_T \quad (5.10)$$

An interesting point regarding this voltage is that if the reference is trimmed so that the output voltage is (5.10), the temperature coefficient at the temperature T_0 is automatically zero. This intuitive result may be difficult to prove in general, but it is actually true for all bandgap reference circuits which generate the full bandgap voltage [110, 78].

We did not have access to detailed process information for V_{G0} and p, but only for the base-emitter voltage of a transistor with a specified geometry at a certain temperature, and a typical temperature coefficient. This is rather typical of the case where the wafer fabricator does not bother to characterize all process parameters, or where these parameters are not controlled. A more empirical method was therefore used to select the component values in the reference circuit.

The process data specifies the base-emitter voltage at a constant bias current I_{C0}, and its temperature coefficient α. The temperature dependence of V_{BE} can then be approximated as

$$\begin{aligned} V_{BE}(T) &= V_{BE0} + \alpha T + T\frac{d}{dT}\left(U_T \ln\left(\frac{I_C(T)}{I_{C0}}\right)\right) \\ &= V_{BE0} + \left(\alpha + \frac{k}{q}\right) T \end{aligned} \quad (5.11)$$

The effective temperature coefficient of the base-emitter voltage, which takes into account the dependence on the bias current, is the expression inside the last brackets. The PTAT current source and the resistor R_2 must therefore be adjusted so the voltage $I_{D5}R_2$ has a temperature coefficient which is the opposite, or

$$\frac{R_2 W_5}{R_1 W_4} \ln(m) = 1 - \frac{\alpha q}{k} \quad (5.12)$$

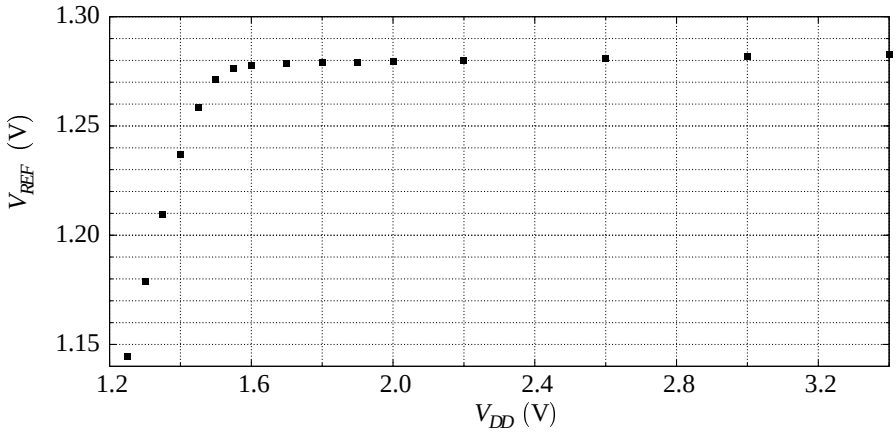


Figure 5.2: *The measured output voltage of the bandgap reference as a function of supply voltage.*

The parameter α is subject to the normal process variations, as are the mirror ratios and the emitter area of the transistor. The layout of the transistor Q_1 was not published in the process data except for the emitter area. This voltage reference was laid out with a substrate transistor of the same emitter area, but the location and size of collector and base contacts were probably not exactly the same as in the published transistor.

The start-up circuit shown in figure 5.1 is not strictly necessary, since a MOS-based PTAT core of this type only has a single stable operating point, whereas a bipolar one has two [116]. The start-up mechanism in a MOS core without an explicit start circuit does however depend on leakage currents in the top current mirror, and these leakage currents can take a very long time to start the circuit. In a system which is designed for intermittent operation or where the functionality must be guaranteed a short time after the power is turned on, the inclusion of a start-up circuit is a wise precaution.

The branch currents in the PTAT core were set to 1 µA, and the mirrored current to the output was set to 2 µA. The temperature coefficient condition in (5.12) was closely fitted with low integer ratios between components for better matching. The transistor width ratio was set to $m = 6$ and the resistor ratio to $R_2/R_1 = 6$, with nominal values of $R_1 = 46.3\,\text{k}\Omega$ and $R_2 = 278\,\text{k}\Omega$. A good feature of this circuit is that the output voltage is independent (to first order) of the resistor values, and only depends on a ratio.

The NMOS transistor dimensions were set to $W_1 = 320\,\mu\text{m}$, $W_2 = 1920\,\mu\text{m}$ and $L_1 = L_2 = 6\,\mu\text{m}$. While these dimensions do not quite fulfill the condition (5.2) at a bias current of 1 µA, they do give an inversion coefficient of 0.06. The rather large gate length was used for better matching of the transistors. The p-channel transistors in the top mirror have very long gates (20 µm) and need no cascodes. The mirror ratio W_5/W_4 was set to 2.

Including the cascode bias circuit and the start-up circuit, the total current consumption of the reference is 4.5 µA.

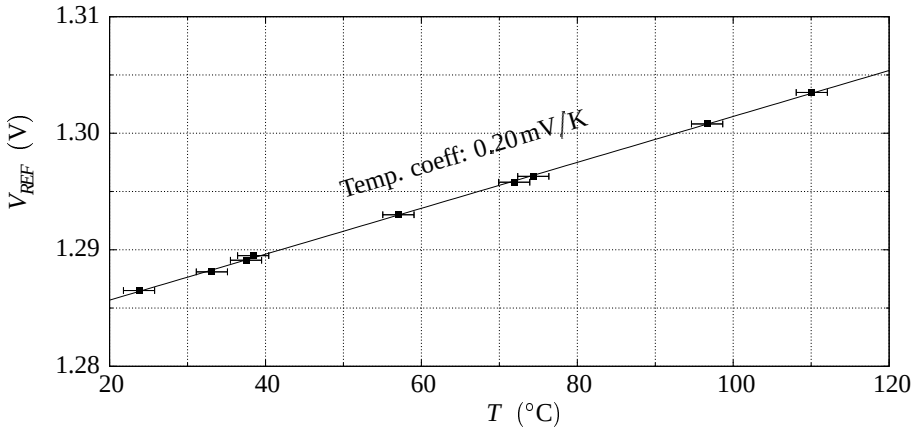


Figure 5.3: *The measured temperature dependence of the bandgap reference voltage and the best linear fit.*

5.1.3 Measurement results

There were 10 packaged samples of the reference, with measured output voltages in the range 1.26–1.28 V at 300K. While this is somewhat larger than the expected value of 1.25 V, the error is not surprising. According to [42], one should not expect an untrimmed accuracy better than 5% from a bandgap reference in an uncalibrated process.

The dependence of the output voltage on the supply voltage was measured (at DC) and it is shown in figure 5.2. The output voltage is within 1% of the final value at a supply voltage of 1.5 V, and the measured slope of the output voltage is 3.1×10^{-3}. The slope is mostly due to the output conductance of the p-channel transistors M3–M5, and could have been improved by adding cascodes to them. This was however not deemed necessary.

The temperature response of one sample was measured by placing it in an oven and recording the output voltage. The voltage was compared to a commercial bandgap reference circuit (Linear Technology's LT-1004-1.2) for better accuracy, using a differential measurement setup. The results, which were rather disappointing, are shown in figure 5.3. The straight line is a linear fit to the data, and the curve just seems to rise linearly, with no maximum in sight. The positive PTAT part of the reference apparently dominates in relation to the negative temperature coefficient of the base-emitter voltage. This is actually not surprising since an output voltage which is above the target value at T_0 leads to a positive temperature coefficient. It would have been worthwhile to confirm this experimentally by verifying that one of the chips with a slightly lower output voltage at T_0 had a lower temperature coefficient, but alas, this was not possible. The homemade heater/oven contraption which was used for these measurements burst into flames at the end of the first run.

The measured temperature coefficient of the reference was 0.20mV/K, about 10 times smaller than the coefficients of the V_{BE} or PTAT source by themselves. The compensation mechanism was therefore working, but it was not eliminating the temperature dependence completely, due either to processing variations in the bipolar transistor or mismatches in the PTAT core or the R_2/R_1 ratio. Had it been possible to trim the resistors, either to increase

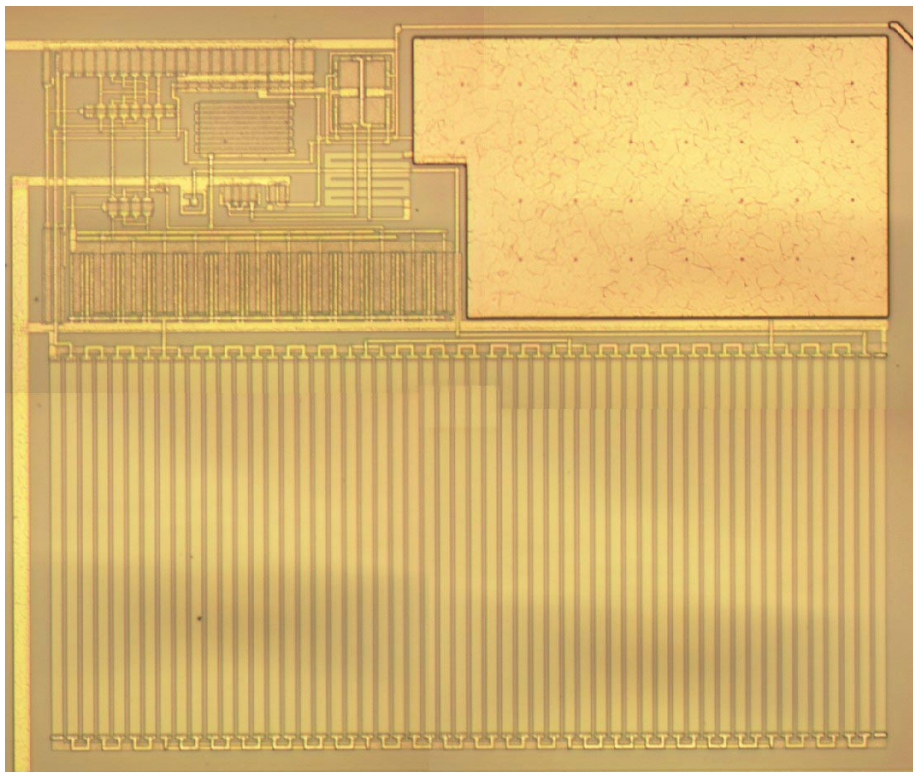


Figure 5.4: *A microphotograph of a bandgap reference similar to the one discussed in the text. This is a different design which includes an opamp buffer, but they have many common features like the matched resistor and transistor arrays.*

R_1 or decrease R_2, the temperature coefficient could have been adjusted to zero at T_0.

5.1.4 Output resistance

The output resistance of the reference in figure 5.1 is $R_2 + U_T/I_{D5}$, which is approximately 300 kΩ. A general expression for the output resistance can be found by using (5.3) and (5.9) with the bias current in the output branch as the free variable:

$$R_{OUT} = \frac{1}{I_{D5}} \left(V_{G0} - V_{BE0} + pU_T \right) \tag{5.13}$$

A lower output resistance can clearly be achieved by spending more current in the output stage but a value much lower than 100 kΩ can not be reached with micropower current levels. The reference can be used as it is to drive capacitive loads but the output should be buffered for heavy resistive loads.

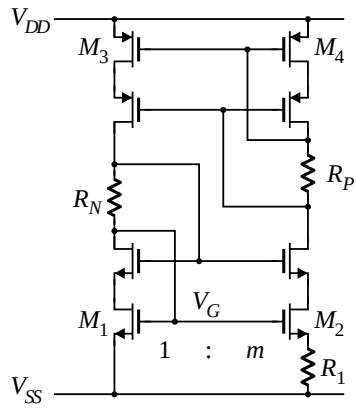


Figure 5.5: *A bias cell version using low-voltage cascodes which do not need separate cascode bias circuits. The start-up circuit is not shown.*

5.2 Bias signal generation

Analog circuit blocks invariably need either a bias voltage or current to generate the correct branch currents and/or node voltages. The bias signals for this purpose can be generated externally, as in the standard solution where an external current generator is used to drive an on-chip bias cell, which then provides bias signals for the chip. Such a setup is useful in a test circuit or a production system where the additional external components can be tolerated. In an implanted system however, these should be eliminated if possible. The reason is not only that the external components and interconnections increase the bulk of the system and the failure probability, but also that any off-chip circuit loop can pick up induced voltages from the strong RF electromagnetic field. On-chip generation of bias signals is presented below, along with an analysis of a specific performance limit for a popular bias cell.

5.2.1 Bias cells

The standard solution for on-chip biasing is to use the so-called constant-g_m cell which is essentially the PTAT core used in the bandgap reference (figure 5.1). The name refers to the fact that the transconductance of M1 is almost independent of temperature. This can be a useful feature, for example for biasing amplifiers whose GBW product is of the type g_m/C.

The basic cell with four transistors and one resistor can be modified in various ways, for example by adding opamps to increase the accuracy and to reduce the temperature dependence [63]. It is also possible to insert any device with a nonlinear I-V characteristic in series with the source of M1 to change the condition for the equilibrium current. The current is then determined by the equality of the voltage drops over R1 and the new device, assuming $W_1 = W_2$. This kind of modification is however not useful for micropower bias circuits where we desire bias currents on the order of 1 µA. Since the voltage drop over R1 in the basic constant-g_m cell is only $U_T \ln(m)$, one can achieve small bias currents with modest values of R1. If the voltage drop were equal to a junction forward voltage drop, as would be the case if one were to add a p-n junction in series with the source of M1, the resistance

would have to be prohibitively large for the same bias current.

In the bandgap PTAT cell, a separate bias circuit was used for the low-voltage cascode on top of the n-channel transistors. This is not strictly necessary since it is possible to obtain the same performance with a self-biased low-voltage cascode as shown in figure 5.5. The supply voltage needed by this bias cell is however larger by two saturation voltages than that needed by the cascode-free version. This circuit may therefore not be practical in low-voltage designs. The insensitivity to supply voltage variations afforded by the use of cascodes in the bias cell may not be required, and this is something which must be decided based on the specifics of the system.

The cascode bias resistors in figure 5.5 should be chosen so the voltage drop over them is equal to the threshold voltages of the cascode transistors. At low bias currents, this can require impractically large resistances, so the resistors may in that case be replaced by active devices [20].

5.2.2 Inversion limits for degenerate mirrors

The correct operation of this bias cell depends on the weak inversion of transistors M1 and M2. We have already mentioned the inversion criteria in (5.2) and the requirement that the inversion coefficient should be much smaller than 1. These are however not very useful for the design of a bias circuit or a PTAT core, since they do not say anything about the error caused by using anything less than perfect weak inversion.

The analysis of transistors in weak inversion is relatively simple since it tends towards the exponential characteristic given by (4.45). Weak inversion is however an ideal case, and if the transistors are to have reasonable dimensions, they must be biased somewhere in the region between weak and moderate inversion. An accurate analysis of the degenerate current mirror must therefore take into account moderate inversion effects. The inversion coefficient $ic = I_D/I_S$ which was defined before, with the specific current defined by

$$I_S = 2n\frac{W}{L}\mu_n C_{ox} U_T^2 \tag{5.14}$$

is a useful parameter. Following the treatment in [34], the normalized drain current of a MOS transistor consists of the forward and reverse currents given by

$$i_D = \frac{I_D}{I_S} = i_F - i_R = F(v_P - v_S) - F(v_P - v_D) \tag{5.15}$$

where v_P is the normalized pinch-off voltage approximated by

$$v_P = \frac{V_G - V_{T0}}{nU_T} \tag{5.16}$$

and the normalized drain and source voltages are $v_D = V_D/U_T$ and $v_S = V_S/U_T$. The second F term in (5.15) can be ignored if the transistors are in saturation, and the inversion coefficient is therefore equal to $F(v_P - v_S)$. The approximating function $F(x)$ is an empirical interpolation between the weak and strong inversion functions, and is not based on any physical effect. The version used here, which gives quite accurate results, is

$$F(x) = \left[\ln\left(1 + \exp(x/2)\right)\right]^2 \tag{5.17}$$

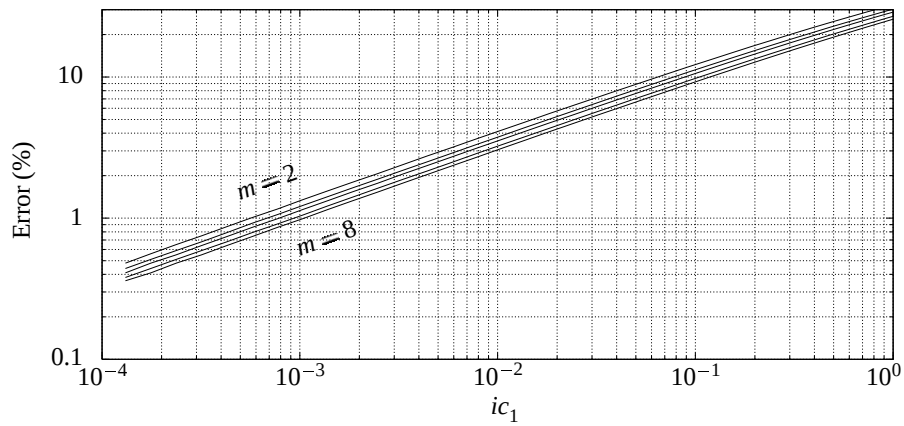


Figure 5.6: *The relative error in the bias cell branch currents caused by the weak inversion approximation. The error is shown as a function of the inversion coefficient of M1, for the current mirror ratios* $m = 2, 3, 4, 6, 8$.

We have applied this formulation of the drain current to transistors M1 and M2 in figure 5.5, assuming that both were in saturation. The drain currents are

$$I_{D1} = I_{S1} F\left(\frac{V_G - V_{T0}}{nU_T}\right) \tag{5.18}$$

and

$$\begin{aligned} I_{D2} &= I_{S2} F\left(\frac{V_G - V_{T0}}{nU_T} - \frac{V_{S2}}{U_T}\right) \\ &= I_{S2} F\left(\frac{V_G - V_{T0}}{nU_T} - \frac{I_{D2}R_1}{U_T}\right) \end{aligned} \tag{5.19}$$

The last equation does not define the drain current of M2 explicitly, and must be solved numerically. These equations were used to obtain an estimate of the error committed by using an inversion coefficient larger than zero in the two transistors. The procedure was to select a gate voltage V_G and calculate the resulting drain current in M1 from (5.18). The drain current in M2 was then set equal to this, and the degenerating resistor R_1 necessary to give that drain current in M2 was calculated through a numerical solution of (5.19). This resistance value was then compared to the ideal value given by (5.1). The ratio between the two values is the relative error in the drain current of M2 caused by moderate inversion effects. The relative error is plotted against the inversion coefficient of M1 in figure 5.6, which shows that the real current is always larger than the asymptotic weak inversion behavior indicates. The figure also shows that a very low inversion coefficient is necessary for a relative error of 1%, and that the error depends to a small extent on the mirror ratio m.

One nice feature of these results is that they do not depend on any circuit parameters such as the transistor width, the gate length or the gain constant $\mu_n C_{ox}$, except through the inversion coefficient.

5.3 Reset circuits

As we have explained in the introduction, the power-on phase of an implanted system must be controlled in such a way that the digital logic blocks wake up in a known state. In the absence of an externally supplied reset signal, the system must sense the operating conditions correctly and generate an internal reset signal at the appropriate time. The specification of the reset operation depends on the nature of the processing elements contained within the system, but a wide range of combinations can be served by a relatively simple reset scheme.

5.3.1 Reset strategy

The pure analog elements should if possible be self-starting so that they do not depend on an external reset signal. Otherwise there can arise circular dependencies which prevent the system from starting up into a correct state. As an example one can imagine a bandgap reference which needs an external reset signal, used alongside a reset circuit which compares the supply voltage to the bandgap reference output. This example is nearly trivially simple, but more complex interdependencies can exist which are not as obvious. One solution is to make all primary analog circuits self-contained. The "primary" tag denotes here all circuits which do not enter into signal paths directly, but provide the necessary operating conditions for blocks in the signal path. Some examples of primary circuits are bias generators, voltage references, supply conditioning/regulator circuits and of course the reset signal generator itself.

Mixed-mode and pure digital blocks which contain sequential logic must however be reset into known state at power-on. Commercial microprocessors and microcontrollers typically use a reset scheme with the reset signal becoming active when the supply voltage falls below a specified threshold. The active-to-inactive reset transition is usually delayed with respect to the change in the controlling condition, to avoid potentially unstable feedback situations. The delay can be set by a simple analog delay element [96] or by a digital timer/counter [81]. Some measure of hysteresis around the toggle point is also desirable for increased stability, so most commercial solutions provide about 20–50 mV of hysteresis.

The state elements in pure digital circuits should be designed with reset inputs so they can be reset into an initial state. Synchronous reset inputs on state elements are in most cases better suited to low-power designs because asynchronous resets involve pulling internal nodes to one of the supplies. This leads to nonzero static power dissipation in CMOS logic families, and increased power dissipation in other types of logic. A synchronous reset strategy does however require at least one active clock edge while the reset signal is active, which places some demands on the design of the clock generator circuit. If the clock generator is inside the system, as is the case with implanted devices, the clock generator is one of the primary reference elements in the system. In other words, the clock generator should be on line at a lower supply voltage than the reset threshold. The clock frequency in implanted systems is normally derived either from a crystal oscillator [5, 98, 51] or from a telemetry carrier frequency [4, 50, 58, 66, 92]. Requiring those to be functional and within frequency specifications during the reset can push the reset threshold to unnecessarily high levels. The correct operating frequency is however not necessary, since only one clock edge is required for the reset. As long as there is a local oscillator in the clock generator which oscillates at a minimum frequency there should be enough clock edges to reset the state of the system.

Crystal oscillators can be designed for operation at 1.1–1.4 V or lower [115, 79, 112], with special processes allowing operation down to 0.45 V [105]. A reset threshold of around 1.5 V is therefore sufficient with regard to the clock signal. In case the system clock is derived from an RF carrier, there is usually a VCO inside a PLL from which the clock is derived. Due to the low frequency range and relaxed spectral purity requirements of inductive links, the VCOs are either relaxation types or ring oscillators [22, 21, 58, 4]. While the PLL may not be locked onto the carrier frequency at low supply voltages, and the clock frequency therefore not equal to the nominal value, the VCOs can easily be designed to operate with supply voltages in the 1 V range.

With a synchronous global reset, the reset threshold level is then governed by two parameters: the supply voltage necessary for the clock generator and the supply voltage necessary for the logic elements. If the clock generator and the logic are supplied by separate supply regulators, the characteristics of the regulators must be taken into account. Standard CMOS logic has several attractive properties which make it a good choice for low-power implanted systems. The static power dissipation is zero, so placing a block in a low-power standby mode is just a question of adding an enable signal. The layout is compact, although fully static memory elements are relatively large. The main disadvantage of standard CMOS logic in a mixed-mode system is the presence of very large supply current transients which can degrade the analog signal quality. The noise problem can be reduced by buffering the supply (section 4.3) or by using low-noise logic families like current steering logic [71, 90, 89, 64]. Low-noise logic types frequently require more power than standard CMOS, so if the digital processing is prominent in the system's power budget the logic types may be partitioned by physical location or by temporal activity. In any case, the necessary supply voltage for standard CMOS, CSL and most other logic types is equal to a single MOS threshold voltage and one or two saturation voltages. With a modern CMOS process, this places the minimum supply voltage below 1 V.

Adding a counter to the reset circuit to provide a controlled time delay is a nontrivial undertaking since the counter elements must have their own internal initialization and some sort of internal or external clock signal must be provided. It can be done by using a two-threshold supply voltage sensor, a clock generator and a counter with an asynchronous reset as shown in figure 5.7. This complexity level is however unnecessary in most systems, where the simple scheme outlined above suffices.

5.3.2 An effective low-power reset circuit

The reset circuit proposed here uses nothing but hysteresis to eliminate the possibility of oscillations in the reset signal as the supply voltage passes through the reset threshold. Since the lack of time delay (except for the delay inherent in a slow, low-power circuit) can increase the tendency of the circuit to produce glitches on the reset output, we have built in a generous hysteresis margin. For a nominal supply voltage of 3 V, the hysteresis of the circuit is set to approximately 0.25 V.

The demands upon the internal reset signal generator listed above have led to the development of the circuit shown in figure 5.8. The supply voltage sensing is based on a comparison of two different functions of the supply voltage, namely the drain currents of M3 and M8. The toggle point of the reset signal is set equal to the supply level at which the currents are equal to each other. The required hysteresis is provided by M5 and M7, but they

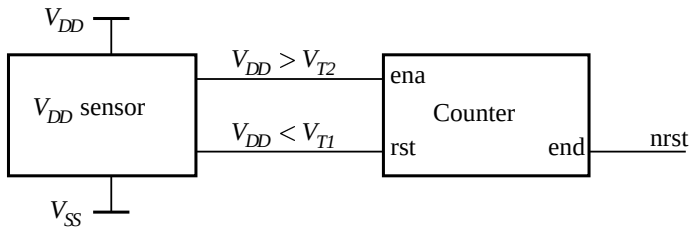


Figure 5.7: *A two-stage reset circuit which guarantees a delay between the time at which the supply exceeds the threshold V_{T2} and the time at which the reset signal becomes inactive. The first threshold V_{T1} is used to generate the asynchronous reset for the internal counter ($V_{T2} > V_{T1}$). The first threshold controls the asynchronous reset input of the counter. The clock generator driving the counter may be internal or external, but it must be running by the time V_{DD} reaches V_{T2}.*

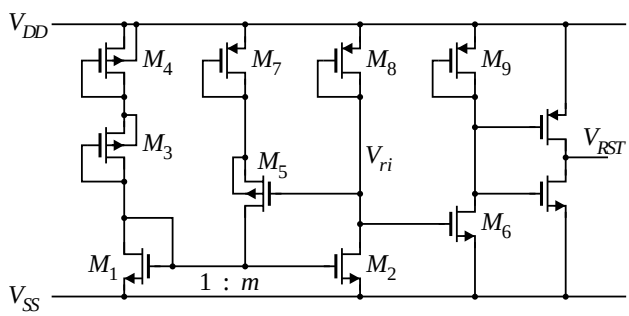


Figure 5.8: *The proposed autonomous power-on reset signal generator. All bulk connections are to the supplies unless the figure indicates otherwise. Relevant transistor sizes are $W/L = 1.2/10$ for M3 and M4, 1.2/150 for M8 and M9, 1.2/60 for M7 and $m = 10$. All transistor dimensions in micrometers.*

can be ignored in the first round of analysis. The purpose of M6–M9 and the CMOS inverter on the output is to amplify the internal reset voltage V_{ri} to provide cleaner transitions on the external reset signal.

M3 and M4 have a (comparatively) large width/length ratio, while M8 is very long and narrow. Neglecting the finer points of moderate and weak inversion operation and of channel-length modulation, the current through the M4–M3–M1 branch is

$$I_{D1} = \left[\frac{\sqrt{\beta_p \beta_n}(V_{DD} - V_{Tn} + V_{Tp})}{2\sqrt{\beta_n} + \sqrt{\beta_p}} \right]^2 \tag{5.20}$$

for supply voltages larger than the sum of the absolute values of the threshold voltages. Transistor M2 can be considered as a switch, so the current through M8 is either determined by the full supply voltage or it is zero. The fact that there are three threshold voltages in the M4–M1 branch means that the current lags behind the current through M8 for low supply voltages. As soon as the supply exceeds the sum of the $|V_T|$'s, the current rises much more rapidly than the drain current of M8. This effect is compounded by the amplification of the current mirror M1–M2. The currents in the two branches are compared in figure 5.9.

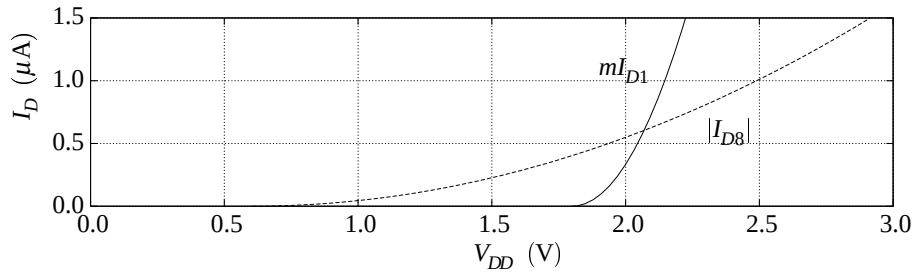


Figure 5.9: *The two currents which determine the toggle point of the reset circuit.*

This simplistic image implies that the current is zero below the threshold, which is not quite true. The NMOS transistors especially, because of their larger aspect ratios and gain factors, operate at a lower gate-source bias than implied by the strong inversion model. The toggle point of the circuit therefore occurs at a lower supply voltage than the intersection point in figure 5.9. The exact value is difficult to calculate analytically, and it is not of enough interest to spend much effort on it. Simulations are adequate for the estimation of the threshold, with process variations probably being a larger contributing factor in the error than simulation inaccuracy. With the dimensions in figure 5.8 the power-on reset signal remains high until a supply voltage of 1.3–1.4 V is reached. This is more than adequate to allow standard CMOS logic circuits to settle into a stable state (assuming threshold voltages of 0.5–0.6 V). The reset toggle voltage scales with the threshold voltages so the circuit is robust with respect to process variations and technology changes. The toggle voltage can be increased by adding more series transistors in the M4–M1 branch.

The hysteresis of the circuit is governed by M5 and M7. The former acts as a switch and the latter bypasses M4 and M3 when the switch is turned on. The hysteresis circuit has no effect when the reset signal is high and the switch is turned off. When the switch is on, the net effect is to increase the current in M1, and referring to figure 5.9 we see that the intersection voltage or toggle point is reduced. The amount of shift in the intersection point defines the hysteresis voltage. A hysteresis gap of 0.2–0.3 V is enough to avoid glitches on the reset signal.

The supply current of the reset circuit is only 2 µA, which compares favorably with published power-on reset circuits (table 5.1). There is no rise in the supply current to speak of in the middle of the transition region, so there is no reverse interaction between the reset signal and the supply voltage. Since the circuit does not use large capacitors or other area-intensive circuits, the layout is very compact.

5.3.3 Measurements

The circuit in figure 5.8 was implemented in 0.5 µm CMOS with the reset signal routed off-chip for testing. The results of rising and falling supply voltage sweep measurements are shown in figure 5.10. The rising toggle point is 1.33 V in this sample which is close to the simulated value. The hysteresis gap is close to 0.25 V as expected.

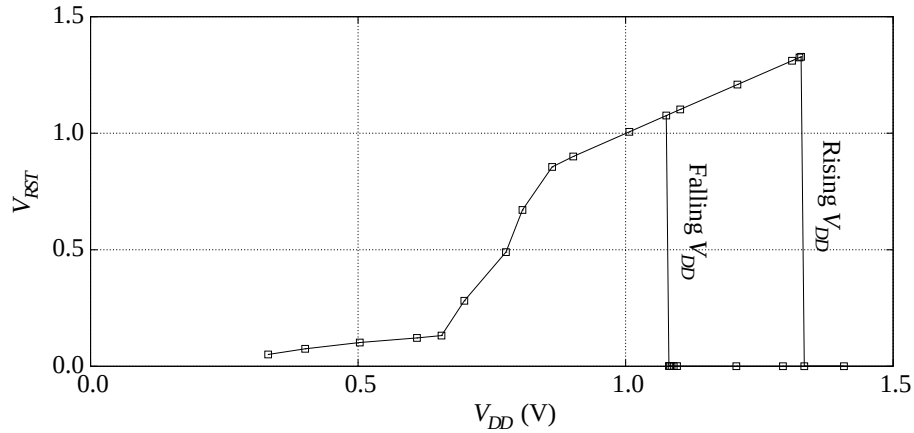


Figure 5.10: *The measured hysteresis curve of the power-on reset signal. The amount of hysteresis is approximately 0.25 V, enough to avoid unnecessary toggling of the reset signal.*

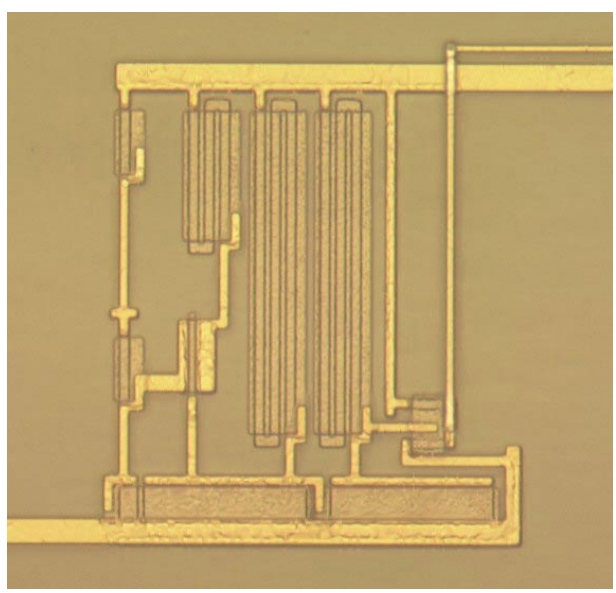


Figure 5.11: *An implementation of a power-on reset circuit very similar to the one dicussed here. A microphotograph of the actual circuit was not available.*

5.3.4 Other circuits

One earlier attempt at designing a power-on reset circuit for an implanted stimulator [47] used a resistive divider to compare the supply voltage to an internal reference voltage, and set the global reset signal according to the outcome of the comparison. While this circuit did function correctly, resetting the system state at power-on, the reset signal was seen to

Circuit	I_{DD} (µA)	V_{hyst}	T_{delay}
This circuit	2	250 mV	NA
LP3470	16	30 mV	Programmable
MAX6381D1–D7	6	NA	1–1800 ms
ADM809	17	0	30 ms
ADM811	5	0	120 ms
ICL7665	3	Progr.	

Table 5.1: *A comparison of the power-on-reset circuit presented here and some recent micropower commercial units. NA means that the information is not applicable or not available.*

toggle repeatedly between the high and low states while the supply voltage swept through a 0.1–0.2 V band around the threshold. This behavior could have been eliminated by the right amount of hysteresis. The inclusion of a resistive divider and a comparator in the reset circuit contributed to the astoundingly high supply current of approximately 100 µA, which is excessive by most modern measures. Some contributing factors to the high supply current were the unusually high supply voltage (12 V), the lack of highly resistive polysilicon and the ancient technology. This example can be used to point out that precision circuit elements like matched resistive dividers and full comparators are rarely needed in a reset circuit. The value of the supply voltage threshold at which the reset voltage toggles can be allowed to vary quite a lot without affecting the operation of the system.

POR circuits are rarely mentioned in the literature, with their correct operation seemingly being taken for granted. A few authors refer to POR circuits without giving details about their construction [122, 4] while others use commercial units in their designs [51, 98]. The power-on-reset circuit presented here is therefore compared to some recent commercial micropower supply monitoring circuits in table 5.1. The commercial devices avoid oscillations on the reset signal in one of two ways: either by applying some hysteresis in the threshold voltages or by adding a dead time after a transition during which the output is fixed.

6
CASE STUDIES

The preceding chapters treat the design of some of the most important building blocks of implantable electronic devices. The integration of these blocks into a complete functioning system has not been addressed in any detail. The goal of this chapter is provide an idea of the limitations and tradeoffs encountered in system design, by reviewing two system IC designs. The former is an implantable stimulator IC which readied in the course of this project. This IC has already been described in detail elsewhere [46, 47] so the focus will be less on the design of the IC itself than on the packaging work done for that chip. The second system design example is a transceiver which is designed as an interface chip between implanted sensor or stimulator ICs and the outside.

6.1 Stimulator chip

This integrated circuit described here was designed for use in an implantable nerve stimulator. It was developed as part of a cooperation between the Department of Information Technology (now Ørsted•DTU) at the Technical University of Denmark and the Center for Sensory-Motor Interaction (SMI) at Aalborg University. This design project, along with most of the other work described here, benefited from the guidance of Dr. Morten Haugland at SMI and his coworkers.

The implantable part of the system consists of the chip itself, several discrete passive components, the antenna coil and a cuff electrode which delivers current pulses to the nerve. The implanted IC is controlled by an external transmitter through an inductive link, with commands consisting of stimulation pulse parameters and stimulation channel number. The cuff electrode has four sets of electrodes and the stimulator has a corresponding number of channels. By use of such a multichannel electrode it is possible to selectively activate specific parts in a nerve trunk [114]. A great deal of work has been done at SMI on the design of stimulators for use by patients suffering from malfunctions of the nervous system. The goal of this design project was to apply integrated circuit techniques to the same problem. A specific condition, called dropfoot, was taken as a reference case. Patients suffering from dropfoot do not have full control over their lower legs, and the idea is to apply a nerve stimulator to the peroneus nerve in the leg to ameliorate the condition.

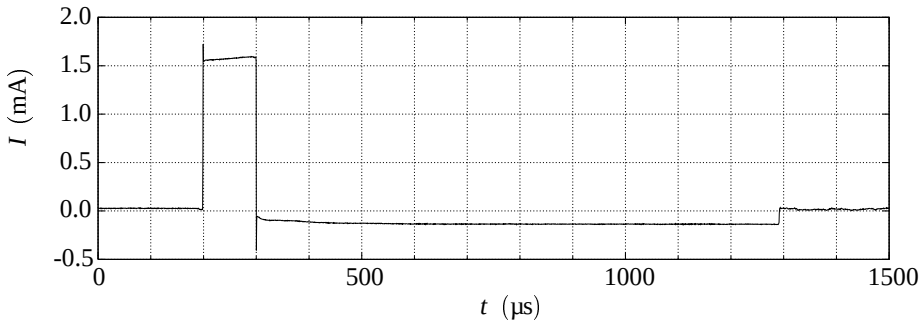


Figure 6.1: *A measured stimulation pulse example. The amplitude of the first pulse phase was set to 1.5 mA and the length to 100 µs. The built-in charge compensation circuit controls the length of the second phase for zero DC current.*

The data transmissions to the implanted system consist of 32-bit command words which are transmitted in bursts, with an unmodulated carrier for maximum power transfer in between. The stimulator reacts instantly to the commands by generating a stimulation pulse, unless a previous command is still active, in which case the new command is ignored. The commands are checked against a transmitted check sum by a 8-bit CRC error detection code [121], and if the check fails, the command is discarded.

The stimulation pulses consist of two phases: the active first phase, which provokes the desired reaction in the nerve, and the second phase whose purpose is to eliminate irreversible electrochemical reactions at the electrode-tissue interface [87, 88]. The amplitude and length of the second current pulse are controlled in order to eliminate any DC current through the nerve electrodes. This has been implemented in other designs either by using DC decoupling capacitors [98, 122] or by controlling the pulse widths for zero total charge [66, 16, 100]. The use of a decoupling capacitor was considered undesirable because of the relatively high space penalty for multichannel stimulators where each output channel must have its own decoupling capacitor. According to our specifications, the charge in each stimulation pulse could be up to 500nC, so the capacitance must be on the order of 500 nF to keep the voltage drop within a reasonable range. Today's high-dielectric constant chip capacitors can provide such capacitances in a 0805 package (approximately $2.0 \times 1.3 \times 1.0$ millimeters) which is quite small, but can still use too much space if many channels are used. The control of the total pulse charge by setting the recharge pulse width depends on the accuracy of the current source and requires a digital calculation of the second pulse width. We were not willing to implement an ALU on the chip to perform the necessary processing, so an analog charge balance method was used. The charge balance method functions by mirroring the stimulation current to an analog integrator, and setting the recharge pulse width to obtain a zero integral. An example of a measured stimulation pulse is shown in figure 6.1.

A microphotograph of the IC is shown in figure 6.2. It contains the receiver described in section 3.2, a phase-locked loop, some logic to process commands and to control stimulation pulses, a voltage regulator for the digital circuits, a DAC for pulse amplitude control, four stimulation current sources, and various support circuits. In was implemented in an old (decrepit, really) 2.4 µm process for the sole reason that it could handle the necessary

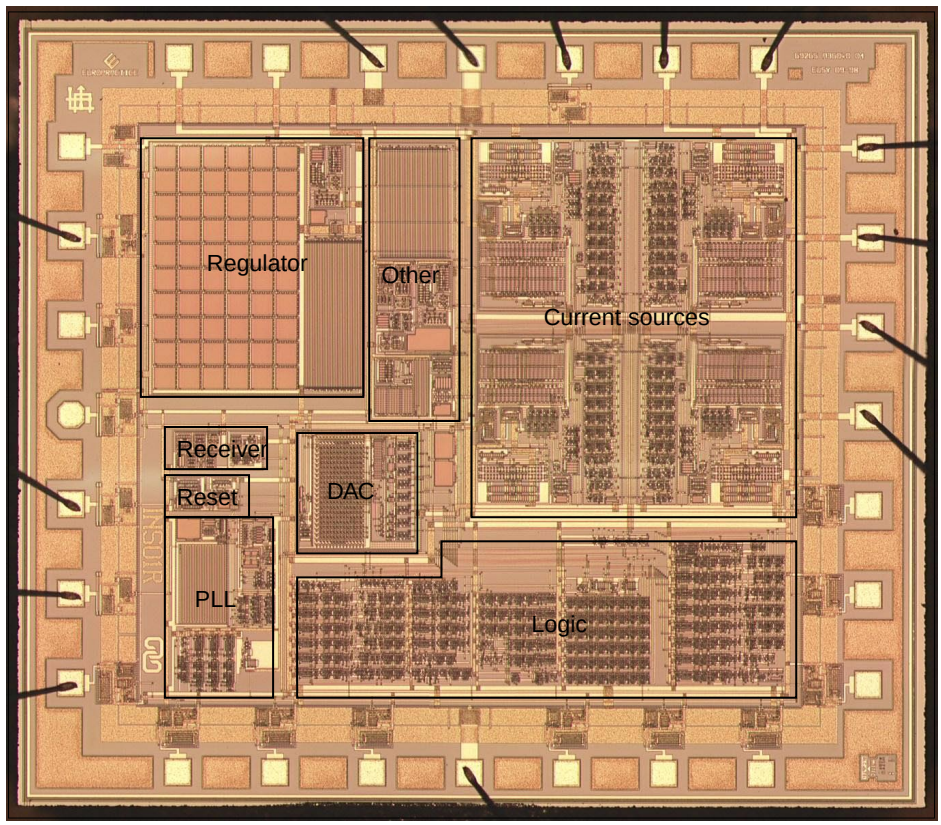


Figure 6.2: A microphotograph of the implantable neural stimulator IC with the main building blocks marked. The chip dimensions are 3.4×3.2 millimeters.

supply voltages for the stimulation pulses. The supply voltage is a recurring problem in the design of neural stimulators. Maximum current pulse amplitudes are often specified at a few milliamperes, and since the electrode impedance can be up to several kΩ, the necessary output voltage range for stimulation current sources is too high for modern IC processes. The only options are therefore to use either an old process as was done here, or to use a modern process with high-voltage capabilities. The latter can however be quite expensive and the design kits appear to be more primitive than for large-volume CMOS processes.

A diagram with all the system components except the electrode is shown in figure 6.3. The diode bridge is available in a single package, but each of the other components corresponds to a single discrete package. The receiver coil L2 was integrated on the substrate. This stimulator IC was implemented before the methods described in chapter 4 were developed, and the overvoltage protection and rectifier bridge were therefore implemented using external components, increasing the bulk and interconnect complexity of the implant. The pulldown resistor R1 was necessary because of an imperfection in the design of the receiver, as described in section 3.2. The PLL loop filter capacitor C3 was necessary because the chip

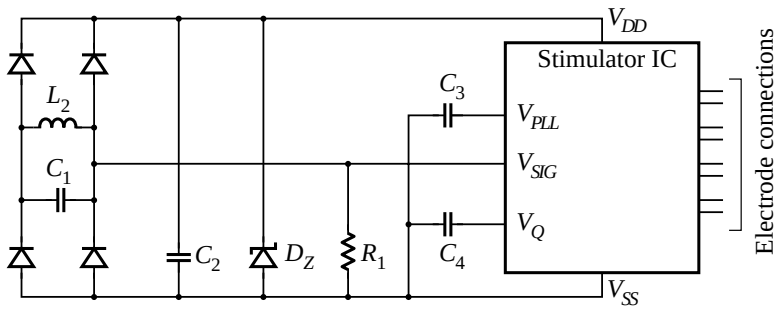


Figure 6.3: *The external discrete components which must be mounted on a substrate with the stimulator chip. They are the tuned LC antenna, the bridge rectifier, a 12 V Zener diode for overvoltage protection, An energy storage capacitor and capacitors for the PLL and charge balance mechanism.*

was designed with an external capacitor to permit testing with different PLL bandwidths. The capacitor could just as well have been included on the chip, and should have been so. The capacitor C4 was necessary for the charge balance mechanism used by the chip.

If a similar stimulator chip were to be designed again using the integration methods in the preceding chapters, together with a time-scaling charge balance method, the only external components would be the charge storage capacitor C2, the receiver coil and the tuning capacitor C1. If the freedom offered by a variable tuning capacitance could be discarded it would be possible also to integrate the tuning capacitor on the IC, since its value is on the order of 100 pF. Switched tuning capacitances like those frequently used in band-switching LC tank VCOs are difficult to implement in this case because of the wide voltage range over the capacitor.

6.1.1 Packaging

A large part of the development work for an implantable system is the development of a packaging solution. Since this aspect of the system design require extensive knowledge of many disparate scientific fields, such as physiology, materials science, biology and chemistry, in addition to electrical engineering, it is not a trivial task. Packaging solutions for the implantable nerve stimulator were investigated as a part of this project, and a prototype was built. This work was carried out at the Microsystems Department at Danish Electronics, Light and Acoustics (DELTA), with guidance from Dr. Haugland.

The outer layers of the encapsulation have the twofold task of protecting the surrounding tissue against harmful materials which may exist in the implant, and to protect the implant itself against the harsh surroundings. Several alternative materials were considered for the encapsulation, for the substrate for the chip and discrete components, and for the cuff electrode. Due to the short time set aside for this encapsulation project, and to our lack of expertise in several critical areas, we ended by copying a packaging concept developed by Dr. Haugland and co-workers. It consists basically of an inner core of epoxy in which the electronic components are embedded, and an outer shell of biocompatible medical grade silicone. The cuff electrode used in the device was also built at SMI. Because we ended up

using mostly an encapsulation solution designed at SMI, we shall not describe it in further detail or take credit for it. We concentrated our efforts instead on the electrical problems in getting the chip to function with the other parts of the system, subject to the limitations posed by the interconnect materials.

The components in figure 6.3 had to be mounted on a common substrate with the chip. We considered for some time using flexible PCB materials, but discarded the idea, as such materials are prone to cracking after long-term implantation [6]. We chose therefore to use a relatively traditional thick-film alumina substrate. This substrate type has also been used at SMI.

The surface of the integrated circuit is protected against the surroundings by a glass passivation layer, except at the bond pads. Since there are high-impedance circuit nodes on the surface of the circuit and the passivation layer is only a few microns, this may be one of the weakest points of the system. Some quantity of moisture and other contaminants must be expected to permeate the encapsulant, since it is not hermetic. To reduce the diffusion of contaminants to the chip surface we chose to flip the chip onto the ceramic substrate. This reduces the area through which contaminants can diffuse in from an area equal to the surface of the chip to an area consisting of a 50 µm wide gap around the periphery of the chip.

A special type of cable was used for connecting the electronics assembly to the electrodes. This is Cooper cable, produced by Finetech Medical Ltd. It consists of four strands of Platinum-Iridium alloy wire embedded in a silicone sheath. Each strand is isolated individually by a polymer coat, and the strands are wound in a helix inside the silicone. The cable is therefore very flexible, and unlikely to cause damage to surrounding tissue. The Pt-Ir alloy is however rather brittle and difficult to handle.

6.1.2 Spatial arrangement

Several subassemblies of the implanted stimulator can be identified, and there are a few different possibilities for the placement of each one. The main subparts are the cuff electrode, the receiver coil and the electronics. They may all be assembled in one compact package and placed directly on the targeted nerve, but several advantages can be obtained by separating some parts from the stimulation site. Three possible configurations are discussed below, each with a specific advantage.

One solution is to place the electronics assembly, the receiver coil and the electrode together at the stimulation site. The relatively compact package obtained in this way has several advantages. The interconnections between the coil, electronics and electrodes can be done locally, inside a common encapsulant, reducing the probability of failure due to mechanical stress or corrosion. The electrical connections between parts of implanted systems have been known to break for a variety of reasons, and a compact package avoids that possibility.

There are however some problems associated with the one-piece implant. The space restrictions at the stimulation site are quite severe, leading to compromises in the design of the implant. An electrode assembly which is too large may crowd the neighboring tissue and ultimately harm the nerve. The receiver coil must therefore be very small, and the low coupling coefficient k in the inductive link leads to excessive power requirements in the transmitter and short battery life. The low k also makes it nearly impossible to obtain data from the implant by passive telemetry methods like load shift keying (LSK).

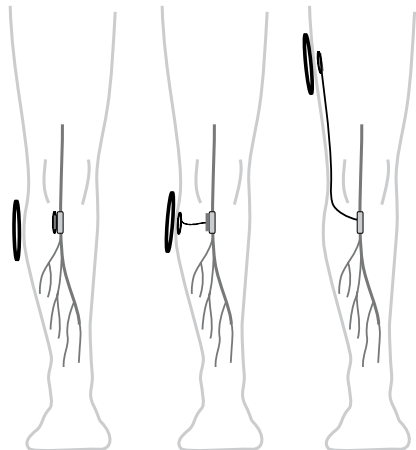


Figure 6.4: *Three possible dropfoot stimulator configurations. The figures show three possible partitioning schemes of the parts of the stimulator, which are the antenna coil, the electronics and the cuff electrode which is mounted on the peroneus nerve.*

The second possibility is shown in figure 6.4(b). In this case, the electronics are still integral with the cuff electrode, but the receiver coil is placed a short distance below the patient's skin, and connected to the implant by a pair of wires. The main advantage of this method is that a good coupling coefficient is achieved between the coils, allowing efficient power transfer to the implant. The bandwidth of the link is also increased, allowing higher data transfer rates and less command overhead. In addition, the increased k allows passive telemetry to be used to extract data from the implant.

The drawback of the method is that the RF signal must be transported some distance over a pair of lossy wires which will degrade the Q of the coil and alter its inductance. The inductance change is however less important than the resistive loss in the wire.

The third possibility is depicted in figure 6.4(c). This resembles a configuration which has been tested extensively at SMI and used in trials with a human patient. A cuff electrode is attached to the peroneus nerve, and is connected to the electronics by a cable. The electronics and receiver coil are placed together below the skin on the patient's thigh. The implant is placed there rather than near the knee to reduce alignment problems. A transmitter mounted near the knee is apparently likely to shift out of alignment with the implanted coil, because of the user's movement.

There are several advantages to this configuration. The relatively shallow implantation of the electronics assembly permits easy surgical access for repairs or replacement. As in case (b) the coil near the skin surface gives a large coupling coefficient and permits passive telemetry. The main drawback of this arrangement is that the electrode connections must be routed between the two parts of the implant. If all 4 stimulation channels are used, 8 separate connections are necessary between the chip and the electrode. Reducing the number of used channels to 2 would simplify the cable requirements to some degree, since

Cooper cable is produced with 4 conductors, but the reduced functionality of the implant would be unacceptable. One of the main features of the stimulator chip is its increased number of stimulation channels, and abandoning that because of packaging problems is rather counterproductive.

The compact on-cuff solution is the most desirable with respect to the electrical and mechanical integrity of the system. There are no external cables that can be broken because of the leg's movement. The movement of interconnect wires in the implant is also restricted because of the encapsulation material, reducing the risk of failure at the connection points. The electrical protection of the components is also best in the compact solution, again because of the lack of long-distance wiring. The physical size of the stimulator can however be a problem, as the external diameter of the cuff is increased by the electronics assembly.

The use of a remote coil as in figure 6.4(b) does reduce the size of the cuff to some degree, and improves the functionality of the implant by allowing bidirectional data transfer at reasonable power levels. The coil connection to the implant consists of a wire pair, which can conceivably be made of some robust material. The coil connection would have to be flexible and elastic, to protect both the user and the connection points to the cuff. A flexible connection is normally implemented using a coiled wire bundle, so this last requirement may be incompatible with the transmission of RF signals through the connection, even over a short distance.

The third possibility, that of separating the electronics from the cuff, is the most challenging in terms of encapsulation requirements. It does however maximize the functionality of the implant, by providing efficient power transfer and thus longer battery life, increased bandwidth, and the possibility of replacement of the electronics. The routing of the stimulation current to the cuff electrode does complicate matters. The electrical characteristics of the connection are not critical, since the signals have a relatively low bandwidth and the current source outputs can overcome the parasitic impedance of the cable.

The last solution was chosen as the best one, with the electronics assembly separated from the stimulation site, and connected to the cuff electrode by a pair of Cooper cables. The receiver coil was integrated on the substrate using the same thick-film paste as the rest of the circuit.

6.1.3 Substrate design

A ceramic thick-film substrate was chosen for the implant because of the low cost, compatibility with the epoxy encapsulation and good physical stability. In addition, this type of substrate is suitable for flip-chip mounting of integrated circuits. The receiver coil was integrated on the substrate, to eliminate the potentially problematic attachment of a wire-wound coil, and because of the uniformity of the resulting inductances. The repeatability of the area and inductance of coils fabricated in this way are much better than for coils wound with copper wire. The use of substrate tracks for the inductor does however lead to higher resistive losses.

The coil was implemented using both sides of the ceramic substrate, with slightly larger than minimum conductor dimensions. The dimensions were increased both out of consideration to the yield of the process (which proved low enough as it was) and to increase the quality factor of the coil. The substrates were manufactured by Hybrico A/S, a Danish firm specializing in the fabrication of thick-film electronics. A substrate with a standard thickness

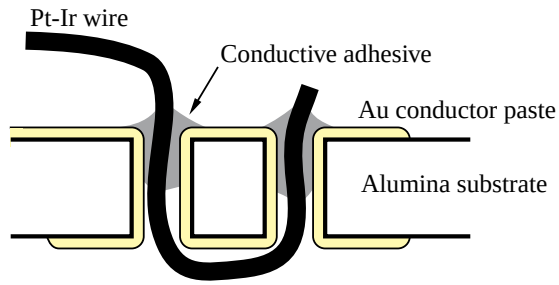


Figure 6.5: *The method used to attach the Cooper cable securely to the substrate.*

of 0.635 mm (1/40th of an inch) was used, with gold paste tracks. A thinner substrate would have reduced the leakage flux between the two inductor layers, but the gain in inductance was not considered worth the cost. Gold tracks were used because of their compatibility with the flip-chip process. The track resistivity reported by the manufacturer is 3–5 mΩ per square. The receiver on the chip is designed for a 5 MHz carrier frequency, and to obtain that resonant frequency with a standard 470 pF capacitor value, the coil inductance was set to 2.15 µH. The field simulation tools discussed in section 2.3 were not available at the time of the design, so a mix of semi-empirical equations and PCB prototypes was used to predict the inductance.

The resistive losses in the receiver coil are due to the bulk resistivity of the material, and at high frequencies, to the skin effect and current crowding. As long as the skin depth

$$\delta = \sqrt{\frac{2}{\mu\omega\sigma}} \tag{6.1}$$

is much smaller than the conductor cross section, it is safe to assume that the resistance is equal to the DC resistance. For a gold conductor, the skin depth at 5MHz is 34 µm. The conductivity σ of the "gold" paste is presumably somewhat lower than for pure gold so the skin depth is even larger. Since the thickness of the track is specified by the manufacturer as approximately 10 µm, the frequency dependence can safely be neglected. The field simulation results in figure 2.7 show that the resistance is constant up to 10 MHz. With the aforementioned track resistivity of 3–5 mΩ per square, the parasitic resistance of the inductor as designed is about 9–14 Ω.

The discrete components were attached to the substrate with conductive adhesive. The mounting procedure for the chip and components was relatively simple, and followed standard practice for attachment of surface mount components. The only special case was the termination of the Cooper cable at the substrate end. A special method was devised for this, by using two vias through the substrate to anchor each wire, as shown in figure 6.5.

6.1.4 Coil fabrication results

Since the inclusion of the receiver coil on the substrate is a novel feature of the system, the characteristics of the fabricated coil deserve to be treated separately. A sample of 20 fabricated coils was subjected to DC resistance measurements and inductance measurements at 5 MHz.

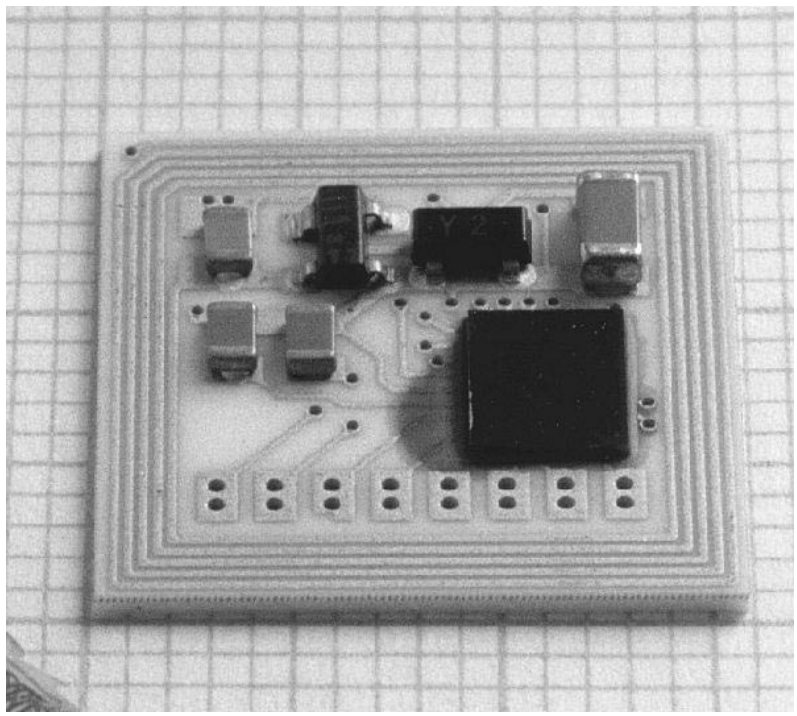


Figure 6.6: *The thick-film substrate with the stimulator and the discrete components mounted on it. The millimeter grid on the background provides the scale. The photo is taken prior to the attachment of the Cooper cable and encapsulation of the substrate in epoxy and silicone.*

Part	Function	Manuf. nr.	Package	Terminals
D_1	Bridge rect.	Siemens BGX50-A	SOT143	SnPb
D_Z	Zener limiter	Fairchild BZX84C12	SOT23	SnPb
C_1	470p tuning cap.	Philips	SMD 0805	AgPd
C_2	470n supply cap.	Philips	SMD 1206	Nickel barrier
C_3	4.7n PLL cap.	Philips	SMD 0805	AgPd
C_4	4.7n Vbal cap.	Philips	SMD 0805	AgPd
R_1	56k Pulldown	Philips	SMD 0805	Nickel barrier

Table 6.1: *Values and manufacturer data for the external discrete components shown in figure 6.3.*

The DC resistance was measured using 4-wire measurements with a low-resistance ohmmeter. Out of the 20 samples, 2 had an infinite resistance, suggesting a break in the track or a contactless via. A microscope was not available at the time of testing, but crude probing with the multimeter suggests that a via was at fault in one of the cases. 16 samples had DC resistances in the range 16.6–17.8Ω, with a mean of 17.1Ω. The two remaining samples had abnormally low resistances of 15.2Ω and 15.6Ω, and visual inspection of the former showed a bridge between two coil turns. It is uncertain whether the latter sample had a fault, or whether it was a case of normal process variations.

The inductances of the 16 "good" samples were measured at 5 MHz with the result $L = 2.5 \pm 0.2\mu$H. This was rather higher than the target value of 2.15 µH, so the inductance estimation methods were obviously not good enough. Later verification with a field simulator (section 2.3) gave values which were within the measurement error margin. Because of the difference between the real and expected inductances, the resonant frequency of the receiver LC circuit was 4.65 MHz instead of 5 MHz. This did not cause any problems, since the transmitter frequency could be adjusted accordingly.

The quality factor of the LC circuit is $Q = \omega L/R_s$, which is approximately 4.2 at 4.65 MHz. While this is lower than the expected value, the coil is still usable, at least to test the implant. The main effect of the reduced Q is to reduce the effective range of the link at a given transmitter power setting.

6.1.5 Measurements

The stimulator assembly was regrettably not subjected to environmental testing, because the assembly and testing work did not reach a final stage within the allotted time span for the project. We did however obtain some results which showed that the IC was functional and responded to commands transmitted over the inductive link.

There were however several basic problems with the system. The overvoltage protection provided by the Zener diode was not quite good enough, both because it was easily destroyed by the overly powerful transmitter and because its current-voltage characteristic was not sharp enough to keep the supply voltage within safe levels. An active solution like the one presented in section 4.1 would have been better in that respect.

The large number of unnecessary external components was also a disadvantage, particularly those which carried sensitive signals, like the PLL filter capacitor and the charge balance capacitor. Because of the intense electromagnetic field, the signals were easily corrupted, and this was probably the cause of the erratic behavior which we observed. Ideally, the only off-chip components should be the supply capacitor and the antenna tuning capacitor.

6.2 Two-chip system transceiver

The discussion in the previous section regarding the placement of the separate parts of the stimulator system emphasized the fact that correct placement of the different system functions is a crucial factor. The antenna, transceiver and power conversion part of the implant should be placed in close proximity to the transmitter to obtain the best link characteristics. However, the strong electromagnetic field from the transmitter is likely to disturb any sensitive signal processing taking place in the implant.

As a solution to this dilemma, we propose therefore to split the implant into two or more units to get the best of both worlds. As a test case, we have implemented a transceiver chip which is intended as a general interface between implanted sensors or actuators and the external world. The general arrangement is shown in figure 6.7, where both a nerve signal sensor and an actuator are shown. The act of dividing the system into parts runs counter to the general principle of using as few components and interconnections as possible, but the advantage of this arrangement is so great that it can be worth the cost. As a reference case, we

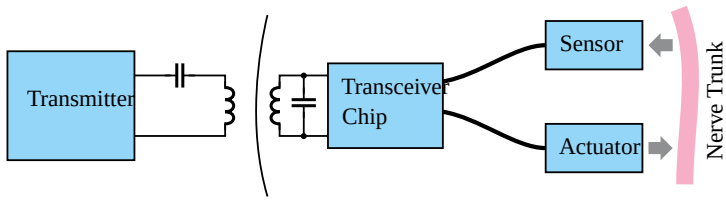


Figure 6.7: *A partitioning example for a FES system containing a transceiver chip and sensor and actuator chips. The last two are not necessarily connected to the same nerve trunk*

use in the following a simple combination of a transceiver chip and a nerve signal amplifier. The amplifier chip can contain the amplifier itself, an analog-to-digital converter and the bus interface logic. By converting the measured signal into the digital domain directly at the source, we avoid the deterioration of the signal which would be the consequence of analog routing over the interconnecting cable.

The transceiver chip contains a considerable amount of digital logic, for control, timing, and buffering of data. The logic will inevitably couple switching noise onto the supplies and into the substrate. Using low-noise logic solutions like current-steering logic [71] can eliminate the problem, but the static supply current of CSL and other ultra-low noise logic families makes them unsuitable as the system complexity passes a certain point. By placing the sensitive analog signal processing on a separate chip, most of the crosstalk from digital to analog is eliminated. A sensor or actuator chip must still contains some logic for buffering and handshake functions, but to a far lesser degree than the transceiver.

Although we speak in the following only of a simple two-chip system, there is in principle nothing against using more than one sensor or actuator with a common transceiver, as shown in figure 6.7. The use of both sensors and actuators in the same implanted system would be a step in the direction of full closed-loop control.

6.2.1 System overview

We adopted a solution with simple digital communication protocols between the central transceiver and control chip, and the outlying transducer chips. The protocol was designed so that it could be accommodated along with the power supply wires in a simple 4-wire Cooper cable. The sensor chip and the transducer chips were actually implemented on the same die. This was mostly done because of the processing convenience, as there were no on-chip connections between the two parts. The sensor chip includes an amplifier for cuff electrode signals, which is itself the subject of an article [72] by its designer, Jannik H. Nielsen. The sensor chip was also supposed to include an AD converter and some bus interface logic, but we were not able to implement them before our deadline. The transceiver includes a direct conversion receiver (actually a homodyne), a load modulation circuit for transmission of data out of the system, control logic, supply voltage conditioning and references.

The following is a description of the transceiver chip and the aspects relating to communication between it and the other parts of the system. Since the main focus of this section is on the way in which partitioning of functions can solve some of the problems in an implanted system of this type, the sensor chip will only be described as far as it relates to the top-level

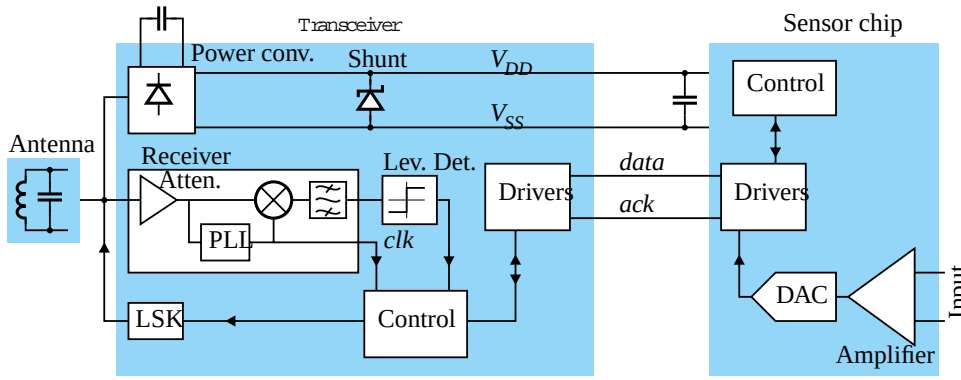


Figure 6.8: *A simplified block diagram of the system, showing the main parts of the transceiver and sensor chips. Differential signals are shown with a single line for clarity.*

design of the system. The principle of partitioning applies to any transducer type that is relevant in an implanted system, so the internals of the sensor chip need not be emphasized. Figure 6.8 shows a top-level diagram of the system.

The chosen modulation method was pulse amplitude modulation (PAM) with Manchester encoding and a modulation index of approximately 0.2. This modulation method is compatible with most transmitter types, and in particular with the transmitter in section 2.4. The carrier frequency was set to fall within the bounds of the 6.78 MHz ISM band, although the frequency was not hard-coded in the receiver. The bit rate was set high enough to control the timing of the sensor chip and to transfer the necessary data.

The power supply management methods used on this chip were in most respects identical to those described in chapter 4. The power conversion circuit was a PMOS bridge rectifier as described in section 4.2, with a total input capacitance of 2 pF. The loading of the antenna by the rectifier is obviously negligible compared to other capacitances in the system, and can be absorbed in the antenna tuning capacitor which is on the order of 100 pF. The total supply current of the transceiver was approximately 120 µA, which could be handled without effort by the conversion circuit. These numbers do not take into account the consumption of the sensor chip, which was also around 100 µA. As we mentioned in section 4.2, the power converter could easily be scaled up to mA current levels in a 0.5 µm technology.

The overvoltage protection is provided by the shunt regulator described in section 4.1. The only external component of the supply conditioning system is the storage capacitor shown in figure 6.8.

The transceiver chip provides the supply voltage for the other chips in the system. Two strands out of four in the Cooper cable have been assigned to V_{SS} and V_{DD}. Because of the helically wound construction of the cable, and the Pt-Ir composition, the impedance is relatively high at all frequencies. To provide a stable supply voltage for the outlying chips, it is therefore necessary to add decoupling capacitors to the supply lines at the transducer ends. A stimulator chip can require relatively large current pulses from the supply while delivering a stimulus.

The receiver part of the transceiver chip is a direct-conversion receiver. Since the input

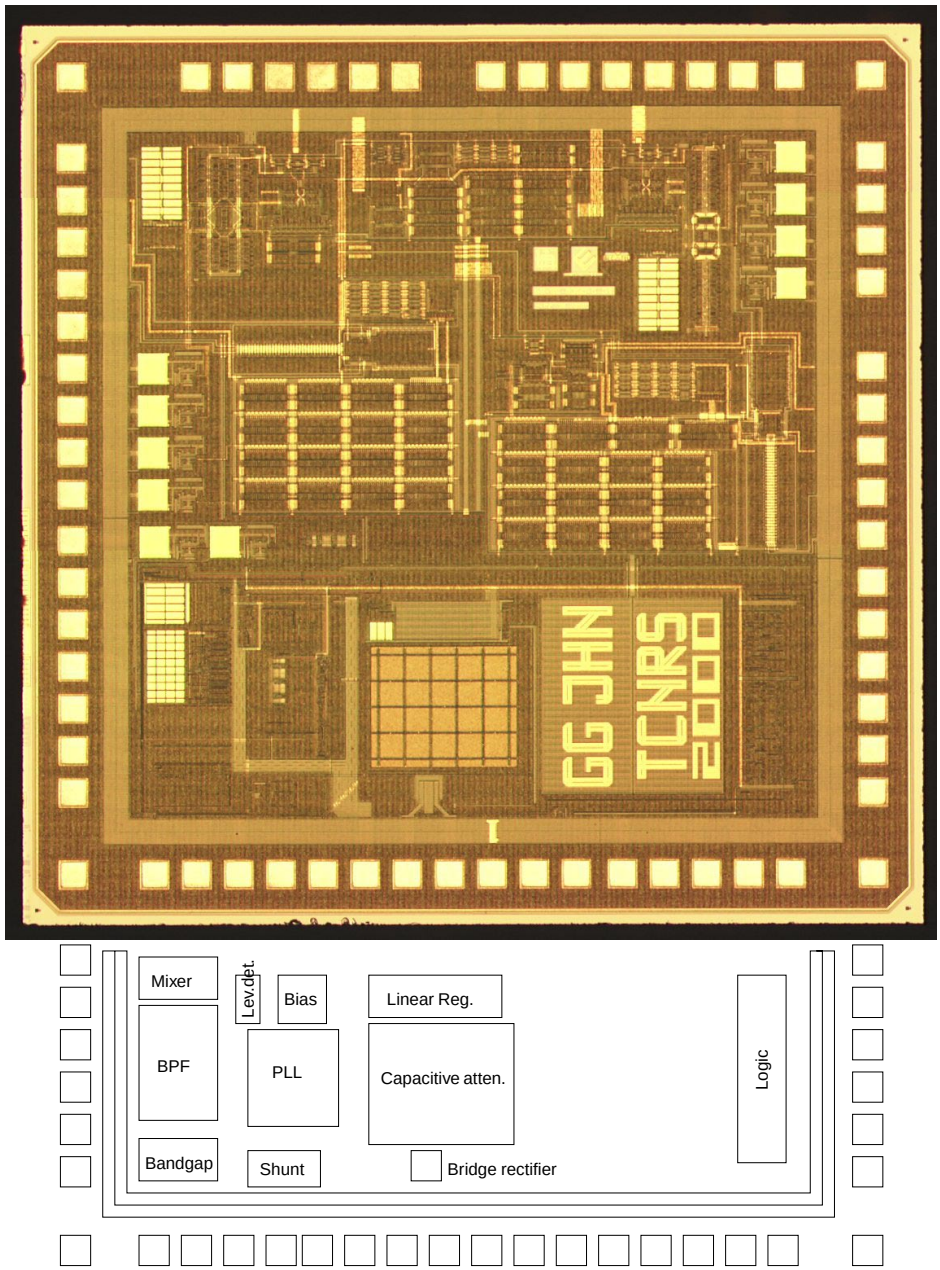


Figure 6.9: The two-chip neural recording system chip. The upper part of the die contains an amplifier for nerve signals designed by Jannik H. Nielsen, described in [72]. The lower part is the transceiver. There are no on-chip connections between the two parts.

signal has a larger amplitude than the supply range, an attenuator with a gain of approximately 0.1 is inserted in the signal path. The input signal is tracked by a PLL which also provides the system clock (the block diagram does not show a clock divider circuit which divides the clock frequency down to 1.7 MHz). The PLL is a traditional charge pump type, with a differential ring oscillator VCO, a phase-frequency detector and a passive loop filter. Due to the large difference in the time constants inside a PLL, it is very inefficient to simulate one using a traditional circuit simulator. In order to simulate the lock time and transient behavior of the loop, we were therefore forced to use other methods. An event-driven simulator was written in C to allow simulation of the PLL, and it is explained in further detail in appendix B.

The output of the PLL and the input signal are put through a mixer, and the mixer output is filtered in a 4th-order differential G_m–C filter. The filter output is directed to a mixed analog-digital level detector, which detects the sign of the transmitted data bits. Instead of using a partly analog level detector, the normal processing method would be to sample the filter output for further digital signal processing. This does however require more sophisticated digital circuits than we were willing to implement, and possibly requires more supply current.

Data is transmitted out of the system by load modulation. The reflected impedance seen by the external transmitter is varied by changing the load seen by the secondary LC circuit. A switch is connected between the terminals of the antenna, and by closing the switch, a maximum change in the reflected impedance is obtained. This simple scheme has the disadvantage of stopping power transfer to the system during load modulation, and other load modulation circuits have been designed to avoid this [102]. The duty cycle of the switch closure is however so low in our case that the reduction in average power transfer is small.

6.2.2 Bus communication

The cable type that we used as a reference for our interconnects is, as mentioned before, a 4-strand biocompatible Cooper cable. Because of the biological constraints that the cable must fulfill, it does not have very good electrical properties, and the use of the cable must be adjusted accordingly. We measured the series resistance at DC of a representative length of the cable, and found it to be 210 Ω/m. The capacitance between any two wires varies from 20–80 pF/m because of the lack of symmetry, and the capacitance of each wire to the surroundings is about 50 pF/m.

No clock signal is sent across the connection between the chips. Instead, we use an asynchronous handshake mechanism to control data transfer, and an internal oscillator in the sensor chip to provide a time base.

The line drivers are class AB circuits which have a quiescent current consumption of 5 µA each, and can slew the line voltages with a 50 µA current. This type of driver was chosen instead of faster types because this is more than sufficient for the purpose, and by limiting the slewing currents the supply transients are reduced. The high and low voltages on the signal lines are 1.0 V and 0.3 V respectively instead of the full supply range, again to reduce supply transients and power consumption.

6.2.3 Results

The basic functions of the transceiver were tested, and data transmission to the chip through an inductive link. The power management functioned without fault, proving that the separate circuits in chapter 4 work well together. Due to time limitations and the lack of some support functions on the sensor chip, we were not able to test the entire system as we would have wished.

With the partitioning scheme discussed here, it is possible to provide better isolation of weak biological signals from strong external disturbances, while simultaneously reducing the overall power consumption by placing the transceiver closer to the external interface.

CONCLUSION

A wide range of solutions for implantable integrated circuit design has been presented in the preceding chapters, solving many of the common difficulties encountered in such designs. The restrictions posed by the implantable nature of the circuits and systems are related to the scarcity of external resources, leading to a need for the integration of all the functionality on a single chip. The resources which are unavailable to the implanted chip are in essence all support functions which are normally taken for granted in chip design, being the supply voltage, biasing signals, voltage references, the time reference. The available power is also severely limited, but that is a limitation which is common to most types of portable electronic devices. In addition to the resource limitations, an implantable IC must also cope with the special mode of operation which mixes data and power transfer in a single signal.

The importance of the inductive link for the system design is emphasized in the first chapter, which presents a short overview of the standard circuit-level description. The standard description is however limited in some ways, since it almost completely ignores the connection between the physical geometry and the circuit representation. It can therefore not be used to predict the range of a link, nor to predict the dependence of the power transfer on the distance. The difficulty lies in the complex dependence of mutual and self-inductances on geometrical details. We have however developed an expression, based on geometrical approximations, which can predict quite accurately the distance dependence of operational parameters, and these expressions have been verified experimentally.

The special method used for inbound data transmission, which is basically modulation of the power transmission, places some unusual demands on the receiver circuits. The additional requirement of having a self-contained receiver, which does not rely on digital signal processing, limits the range of available receiver types. We have developed a completely self-contained receiver for amplitude shift keying (ASK) modulation which is compatible with these requirements, while using a very low supply current. The receiver can handle variable modulation depths, and spans the modulation range of the most efficient transmitter architectures. Where previous designs have relied on preset modulation levels or constant modulation of the carrier to set the detector levels, our design does away with such limitations. The receiver operates over a wide range of carrier frequencies and bit rates, corresponding to the prevalent values used in inductively coupled systems, and can without difficulty be modified to extend this range.

The power supply management methods are treated separately, as many of the special conditions for implanted ICs have a large impact on the power supply quality. We have developed and implemented a two-terminal overvoltage protection circuit, or a shunt regulator

in other words, which is completely self-sufficient. This circuit can be used instead of the commonly used passive protection circuits, internal or external, and is compatible with plain CMOS processing. The knee voltage of the protection circuit is accurate to within a few percent, due to the use of an integrated bandgap voltage reference. The protection circuit spans a very wide current range, and can protect the IC against the worst power transients without drawing a noticeable current when it is not needed. The stability criteria of the shunt regulator were derived, along with an expression for the maximum shunt current range. An implementation of the protection circuit confirmed their validity, and performed according to expectations, showing that this solution is superior to those used in previous designs.

The rectification of the carrier for the conversion to a DC supply voltage was addressed, with the objective being complete integration in a CMOS technology. A bridge rectifier implementation was developed which avoids the pitfalls of loss currents stemming from the CMOS parasitic elements. The loss currents were reduced by a large factor by using p-channel MOS transistors in the rectifier bridge and biasing them in weak inversion. The speed penalty normally associated with weak inversion operation is not relevant in this case, since the modest increase in capacitance can be absorbed in the antenna tuning capacitor. The ratio of the desired rectifier current to the parasitic loss current was derived, and it was shown that a very large current ratio could be attained, effectively reducing the losses to zero. The dependence of the improvement on layout factors was also examined, and an alternative improved layout was suggested.

Any integrated system chip of the type described here will unavoidably consist of a mix of analog and digital circuits. Due to the single shared supply point of the entire system, the digital supply current transients are seen unfiltered by the analog part. We have proposed the use of an on-chip linear regulator, both to provide isolation between the digital and analog circuits and to reduce the supply current of CMOS logic. We have described a regulator design which is based on a previous design, but improves on it by extending its supply voltage range.

On-chip generation of bias signals is almost invariably implemented by using degenerate current mirrors, due to the predictability of the resulting bias currents and the low sensitivity to supply voltage variations. If these bias generators are implemented in CMOS, the degenerate current mirrors must be biased in weak inversion. We have analyzed the impact of moderate inversion effects on the resulting bias current, and present guidelines for the required biasing level to obtain a given maximum deviation from the ideal case.

The reset signal is another example of a general circuit resource which must be generated internally in an implanted system. Different reset strategies for mixed analog/digital systems were discussed, and a new reset signal generator was presented. The reset circuit is based on supply voltage sensing using a comparison of two different functions of the supply voltage. The circuit provides a sufficient amount of hysteresis in the reset signal threshold to avoid false triggering due to supply voltage fluctuations, and uses very little current.

An overview of system design for implantable ICs was provided by the use of two examples. The first example is an implantable neural stimulator chip, whose initial design predates the time frame of this project. The chip was however finished during the course of this study, and it is used to illustrate some of the interface and packaging problems common to implantable systems. The chip was flipped onto an alumina substrate with discrete support components, along with a novel thick-film antenna coil implementation. Despite the unwanted interaction of the electromagnetic field with some of the externally routed sig-

nals, the system assembly did prove to be functional. This chip did not incorporate any of the integration solutions presented in this thesis, but had it done so, the discrete component count would have been far lower and the system would have been more resistant to external disturbances.

The second system chip example is a transceiver which does use the methods presented here to reduce the external component count to the bare minimum. The transceiver IC is conceived as an interface between implanted sensors or actuators and the external (non-implanted) apparatus. The transceiver is placed in close proximity to the surface of the body, where the field from the transmitter is strongest. The high field strength and coupling coefficient permit effective inbound data and power transmission, and allow the use of passive load modulation for outbound transmission. The sensors or actuators are connected to the transceiver chip by a two-wire bus for serial transmission. This bus is implemented physically by a length of Cooper cable, which also carries the supply power to the sensors or actuators. The main point is to perform any signal processing and conversion between the digital and analog domains locally, at the electrode interface, to eliminate the effect of induced signals and noise coupling.

In summary, the main subject of this thesis is the development of CMOS circuit solutions for implant functions which have previously only been possible with advanced processing techniques, or by the use of external components. We present methods to drastically reduce the component count and interconnect complexity of implanted ICs, and present an example of a system IC which incorporates these methods.

A
FIELD SIMULATION SOFTWARE

The field simulation software used to calculate the self-inductance of receiver (and transmitter) coils in section 2.3 was FastHenry [56, 57, 84]. This program gives quite accurate results in magnetic calculations, and can handle arbitrary geometries. It can be used to calculate mutual inductance matrices in n-ports, of which an inductive link with two coils is a special case. There exist other free software packages like ASITIC [73] which may be as accurate and versatile, but FastHenry was suitable for the task.

The input to the program is a description of the circuit geometry written in a custom description language (see example below). This description defines the circuit in terms of rectilinear segments which are connected together at nodes with given spatial coordinates. An arbitrary number of ports can be defined, although we restricted our simulation runs to two-ports. The segments are partitioned internally in discrete elements, and this partitioning is used to simulate skin effects and frequency-dependent parasitic resistances. Curved circuit tracks like those in a circular coil must be discretized into linear segments, and the number of segments and the degree of internal partitioning affects the accuracy of the final outcome. Obviously, better accuracy can be obtained by a using more segments and finer partitioning, at the expense of longer computing times. The computing time in a single run is not relevant since it is completed is a reasonable time on a fast computer, even for relatively complex geometries. We did however perform sweeps over a range of link geometries with thousands of runs in each sweep. The level of necessary accuracy was therefore determined by finding the point where a finer discretization of the geometry did not give significantly different results.

To make it possible to quickly change the geometry of the link, several device generators were written in C which could generate FastHenry circuit descriptions from a set of parameters, like the coil diameter, conductivity, number of turns and so on. These generators were then called from a shell script which carried out a sweep over a specified range of parameters. This provided an effective tool for the estimation of the effect of specific geometry parameters on the link behavior.

Further information about the use of the program is available from [84]. The code listings below give examples of a circuit generator for a circular transmitter/receiver coil and a generator for a two-layer planar thick-film coil.

```
/* ---------------------------------------------------------------*/
/* fhgen.h                                                        */
/* Header for inductor description generators for fasthenry inp.*/
/* (C) 2001, Gunnar Gudnason                                      */
/* ---------------------------------------------------------------*/
/* All dimensions in millimeters                                  */


/* Interesting conductivity data taken from                       */
/* http://www.eddy-current.com/condres.htm:                       */
/* Source Code: 1 - CSNDT                                         */
/*              2 - Eddy Current Testing Manual                   */
/*              3 - NDT Magazine Sept/Oct 1955                    */
/*  RESIST.   COND,/                SOURCE                        */
/*  ohm-m     SIEMENS         % IACS  CODE    MATERIAL            */
/*  1.591E-08 6.287E+07        108.40   1  Silver, Pure           */
/*  1.642E-08 6.090E+07        105.00   2  Silver, Pure           */
/*  1.664E-08 6.009E+07        103.60   1  Copper, Pure           */
/*  1.724E-08 5.800E+07        100.00   2  Copper, Pure           */
/*  2.028E-08 4.930E+07         85.00   1  Copper, Deoxidized     */
/*  2.349E-08 4.257E+07         73.40   1  Gold                   */
/*  2.463E-08 4.060E+07         70.00   2  Gold, Pure             */
/*  2.655E-08 3.767E+07         64.94   1  Aluminum, 99.99%       */
/*  2.826E-08 3.538E+07         61.00   2  Aluminum, Pure         */
/*  Conductivity data for sintered gold paste is uncertain, but */
/*  the quoted track paste resistance from the manufacturer was */
/*  approx. 3-5mOhm/square. With a track thickness of 100um the */
/*  resistivity is then rho=thickness*Rsquare=1.5-2.5e-7 Ohm*m. */
/*  Model the real track with a thickness of 50u and a resist.  */
/*  which fits the measured DC resistance of about 12 Ohm.      */

#define NMAX 4092
#define COND_Cu (double)5.800E+04
#define COND_Ag (double)6.100E+04
#define COND_Au (double)4.100E+04
#define COND_AuTF (double)4.0E+03  /* Au Thick film paste */

int fhgen_helix(double,double,double,double,double,double,double,
                int,int,int,int);
int fhgen_rect2l(double,double,double,double,double,double,
                 double,double,double,double,int,int,int);
```

```
/*---------------------------------------------------------------*/
/* fhgen_main.c                                                  */
/* Test main file for fasthenry coil generators                  */
/* Generates one circular primary and one stimsub secondary at */
/* relative positions defined by the x and z input parameters  */
/* (C) 2001, Gunnar Gudnason                                     */
/*---------------------------------------------------------------*/
#include "fhgen.h"
#include <math.h>
#include <stdio.h>

main(int argc, char *argv[])
  {
  int n1,n2;
  double xcoord,zcoord;
  double freq,R1;
```

```
  if(argc < 4)
    {
    fprintf(stderr,
        "Usage: foo freq xcoord zcoord [Rprimary]\n");
    exit(1);
    }

  sscanf(argv[1],"%lf",&freq);
  sscanf(argv[2],"%lf",&xcoord);
  sscanf(argv[3],"%lf",&zcoord);
  if (argc>4)
    sscanf(argv[4],"%lf",&R1);
  else
    R1=25.0;

  printf("* File generated by fhgen_main.c\n");
  printf("* FastHenry input file with inductor descriptions\n");
  printf(".units mm\n");

  /* Place a helical copper coil at position (x,0,z) */
  printf(".default sigma=%.6le nhinc=%d nwinc=%d\n",
                                                COND_Cu,1,1);
  n1=0;
  n2=fhgen_helix(R1, .5l, .5l, .7l,
                  xcoord, 0.0l, zcoord,
                  (int)5, (int)16, n1, n1 );
  printf(".external n%03d n%03d\n",n1,n2);

  /* Place a rectangular thick-film coil at (0,0,0), with same  */
  /* dimensions as the ins01r stimsub coil                      */
  printf(".default sigma=%.6le nhinc=%d nwinc=%d\n",
                                                COND_AuTF,1,1);
  n1=1000;
  n2=fhgen_rect2l( 12.15l, 11.75l, 0.20l, 0.06l, 0.15l,
               0.0l, 0.0l, 0.0l,
               0.635l, 0.5l,
               (int)5,n1,n1 );
  printf(".external n%03d n%03d\n",n2,n1);
  printf(".freq fmin=%.6le fmax=%.6le ndec=1\n",
                                                freq,freq);
  printf(".end\n");
  }
```

```
/*--------------------------------------------------------------*/
/* fhgen_rect2l.c                                               */
/* Two-layer inductor description generator for fasthenry input.*/
/* Generates a geometrical description of a two-layer planar    */
/* spiral inductor on a substrate for                           */
/* for inductance and resistance estimation.                    */
/* (C) 2001, Gunnar Gudnason                                    */
/*--------------------------------------------------------------*/
/* All dimensions in millimeters                                */

#include "fhgen.h"
```

```
#include <math.h>
#include <stdio.h>

/* Returns the number of the last node                          */
int fhgen_rect2l(
double dix, double diy,             /* Internal diameter x and y*/
double cw, double ch, double sp,    /* Conductor w, h, spacing  */
double x0, double y0, double z0,    /* Coil center              */
double thick,                       /* Substrate thickness      */
double chamfer,                     /* Corner chamfer tangent   */
int turns,
int n0,                             /* Starting node number     */
int e0)                             /* Starting edge number     */
  {
  int nc=0;                                     /* Node counter*/
  int ne=0;                                     /* Edge counter*/
  int tc=0;                                     /* turn counter*/
  int viaflag[NMAX+4];
  int nodes=0;
  int i;
  double cx,cy,cz;

  /* Use via flag to indicate that segment starting             */
  /*in node i is a vertical via segment                         */
  for(i=0;i<NMAX;i++) viaflag[i]=0;

  /* Only allow coil planes parallel to the xy-plane for now.   */
  /* Adjust xy zero coords because the upper left corner of the */
  /* innermost turn was set to (x,y)=(0,0)                      */
  cz=z0+thick/2.0;
  x0-=dix/2.0;
  cx=-dix/2.0;
  y0+=diy/2.0;
  cy=diy/2.0;

  /* Info section */
  printf("* ------------------------------------------------\n");
  printf("* Thick-film 2-sided coil instance by fhgen_rect2l\n");
  printf("* 2-layer planar inductor on ceramic substrate\n");
  printf("* Inductor info: \n");
  printf("* Center (x,y,z)=(%.4lf,%.4lf,%.4lf)\n",x0,y0,z0);
  printf("* dix=%.4lf diy=%.4lf \n",dix,diy);
  printf("* track w=%.4lf sp=%.4lf h=%.4lf\n",cw,sp,ch);
  printf("* sub. thickness=%.4lf\n",thick);
  printf("* %d turns\n",turns);

  printf("* Nodes:\n");
  nc=0;
  /* Top part of inductor */
  for(tc=0;tc<turns;tc++)
      {
      /* TL corner */
      printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
      nc++;
      /* TR corner */
      cx = x0 + dix + tc*(cw+sp);
      printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
      nc++;
      /* BR corner */
      cy = y0 -diy - tc*(cw+sp);
      printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
```

```
    nc++;
    /* BL corner */
    cx = x0 - (double)(tc+1)*(cw + sp);
    printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
    nc++;
    /* 1st node of 45 degree chamfer, 0.414=arctan(22.5 deg) */
    cy = y0 + (double)(tc+1) * chamfer * (cw+sp);
    printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
    nc++;
    cy = y0 + (double)(tc+1)*(cw + sp);
    cx = x0 -(double)(tc+1) * chamfer * (cw+sp);
    }
/* Last edge on top side of the substrate */
cx = x0 - (double)(tc)*(cw + sp);
cy = y0 + (double)(tc-1)*(cw + sp);
printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
viaflag[nc]=1;
nc++;

/* Vertical connecting segment between sides */
cz-=thick;

/* Bottom part of inductor. */
/* Only the chamfer has different coords from the top */
for(tc=turns-1;tc>=0;tc--)
    {
    /* TL corner */
    printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
    nc++;
    /* TR corner */
    cx = x0 + dix + (tc)*(cw+sp);
    printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
    nc++;
    /* BR corner */
    cy = y0 -diy - tc*(cw+sp);
    printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
    nc++;
    /* BL corner */
    cx = x0 - (double)(tc+1)*(cw + sp);
    printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
    nc++;
    /* 1st node of 45 degree chamfer, 0.414=arctan(22.5 deg) */
    cy = y0 + (double)(tc) * chamfer * (cw+sp) - (cw+sp);
    printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
    nc++;
    cy = y0 + (double)(tc)*(cw + sp) - (cw+sp);
    cx = x0 -(double)(tc) * chamfer * (cw+sp) - (cw+sp);
    }

/* Short hop towards the middle */
cx+=1.5*(cw+sp);
printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
nc++;

/* Upwards via back to the top */
cz+=thick;
printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
viaflag[nc-1]=1;
nc++;

nodes=nc;
```

```
  /* Print edge definitions:                        */
  /* Cross section of vias is set to cw times cw */
  printf("\n* Edges:\n");
  for(nc=0;nc<nodes-1;nc++)
    printf("e%03d n%03d n%03d w=%08.4lf h=%08.4lf %s\n",
           nc+e0,nc+n0,nc+1+n0,cw,(viaflag[nc]?cw:ch),
           (viaflag[nc]?"wx=1.0 wy=0.0 wz=0.0":""));
  printf("* ------------------------------------------------\n");
  return(n0+nodes-1);
  }
```

```
/*---------------------------------------------------------------*/
/* fhgen_helix.c                                                 */
/* Helical inductor description generator for fasthenry input.   */
/* For inductance and resistance estimation.                     */
/* Jan 2001, Gunnar Gudnason                                     */
/*---------------------------------------------------------------*/
/* All dimensions in millimeters                                 */

#include "fhgen.h"
#include <math.h>
#include <stdio.h>

#define MATH_PI 3.14159265359

int fhgen_helix(
double radius,                  /* Radius of cylinder on which*/
                                /* the nodes lie.             */
double cw, double ch,           /* Conductor H and W.         */
                                /* H is in the axial direction*/
                                /* and W is radial.           */
double pitch,                   /* Winding pitch in axial dir.*/
double x0, double y0, double z0,  /* Coil center              */
int turns,
int vertx,                      /* Cross section vertices     */
int n0,                         /* Starting node number       */
int e0)                         /* Starting edge number       */
  {
  int nc=0;                                   /* Node counter*/
  int ne=0;                                   /* Edge counter*/
  int tc=0;                                   /* turn counter*/
  int nodes=0;
  int i;
  double cx,cy,cz;
  double theta,dtheta;
  double dz;

  /* Info section */
  printf("* ------------------------------------------------\n");
  printf("* Helical coil instance by fhgen_helix\n");
  printf("* Inductor info: \n");
  printf("* Center at (x,y,z)=(%.4lf,%.4lf,%.4lf)\n",x0,y0,z0);
  printf("* Helix radius=%.4lf\n",radius);
  printf("* Polygon vertices=%d\n",vertx);
  printf("* track w=%.4lf h=%.4lf pitch=%.4lf\n",cw,ch,pitch);
  printf("* %d turns\n",turns);

  /* Only allow coil planes parallel to the xy-plane for now.   */
  /* Starting point:                                            */
```

```
  cz=z0-((double)turns*pitch)/2.0;
  cx=x0+radius;
  cy=y0;
  dtheta=2.*MATH_PI/vertx;
  theta=0;
  dz=pitch/vertx;

  printf("* Nodes:\n");
  for(nc=0;nc<turns*vertx;nc++)
      {
      printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",
                                              nc+n0,cx,cy,cz);
      theta+=dtheta;
      cz+=dz;
      cx=x0+radius*cos(theta);
      cy=y0+radius*sin(theta);
      }
  /* Last node */
  printf("n%03d x=%8.4lf y=%8.4lf z=%8.4lf\n",nc+n0,cx,cy,cz);
  nodes=nc;

  printf("\n* Edges:\n");
  for(nc=0;nc<nodes;nc++)
    printf("e%03d n%03d n%03d w=%08.4lf h=%08.4lf\n",
           nc+e0,nc+n0,nc+1+n0,cw,ch);
  printf("* ------------------------------------------------\n");
  return(n0+nodes);
  }
```

The description generated by these routines includes material definitions, circuit node coordinates and the properties of the edges in the circuit. The abbreviated output file example below shows the format of the generated circuit description.

```
* File generated by fhgen_main.c
* FastHenry input file with inductor descriptions
.units mm
.default sigma=5.800000e+04 nhinc=4 nwinc=5
* ------------------------------------------------------
* Helical coil instance by fhgen_helix
* Inductor info:
* Center at (x,y,z)=(0.0000,0.0000,50.0000)
* Helix radius=25.0000
* Polygon vertices=16
* track width=0.5000 height=0.5000 pitch=0.7000
* 5 turns
* Nodes:
n000 x= 25.0000 y=  0.0000 z= 48.2500
n001 x= 23.0970 y=  9.5671 z= 48.2938
n002 x= 17.6777 y= 17.6777 z= 48.3375
n003 x=  9.5671 y= 23.0970 z= 48.3813
n004 x= -0.0000 y= 25.0000 z= 48.4250
n005 x= -9.5671 y= 23.0970 z= 48.4688
n006 x=-17.6777 y= 17.6777 z= 48.5125
n007 x=-23.0970 y=  9.5671 z= 48.5563
n008 x=-25.0000 y= -0.0000 z= 48.6000
n009 x=-23.0970 y= -9.5671 z= 48.6438
```

```
n010 x=-17.6777 y=-17.6777 z= 48.6875
<snip>
n079 x= 23.0970 y= -9.5671 z= 51.7063
n080 x= 25.0000 y=  0.0000 z= 51.7500

* Edges:
e000 n000 n001 w=000.5000 h=000.5000
e001 n001 n002 w=000.5000 h=000.5000
<snip>
e078 n078 n079 w=000.5000 h=000.5000
e079 n079 n080 w=000.5000 h=000.5000
* ------------------------------------------------------
.external n000 n080
.default sigma=4.000000e+03 nhinc=5 nwinc=7
* ------------------------------------------------------
* Rect. thick-film 2-sided coil instance by fhgen_rect2l
* 2-layer spiral planar inductor on ceramic substrate
* Inductor info:
* Center at (x,y,z)=(-6.0750,5.8750,0.0000)
* dix=12.1500 diy=11.7500
* track width=0.2000 space=0.1500 height=0.0600
* sub. thickness=0.6350
* 5 turns
* Nodes:
n1000 x= -6.0750 y=  5.8750 z=  0.3175
n1001 x=  6.0750 y=  5.8750 z=  0.3175
<snip>
n1051 x= -5.9000 y=  5.5250 z= -0.3175
n1052 x= -5.9000 y=  5.5250 z=  0.3175

* Edges:
e1000 n1000 n1001 w=000.2000 h=000.0600
e1001 n1001 n1002 w=000.2000 h=000.0600
<snip>
e1050 n1050 n1051 w=000.2000 h=000.0600
e1051 n1051 n1052 w=000.2000 h=000.2000 wx=1.0 wy=0.0 wz=0.0
* ------------------------------------------------------
.external n1052 n1000
.freq fmin=6.780000e+06 fmax=6.780000e+06 ndec=1
.end
```

For calculations which require a large number of time-consuming FastHenry runs, it is impractical to generate the FastHenry input file manually before each run. The data for figure 2.9 for example, required 1100 calls to the FastHenry program, taking a total time of 36 hours on a HP PA-RISC 8600 workstation. The script example below generates geometry description files, calls the FastHenry simulator, calls a back-end processor to extract the coupling coefficient from the resulting mutual impedance matrix, parses the result and puts it in a data file which can then be plotted directly by Gnuplot. This is actually the script which generated the data for figure 2.9.

```
#!/bin/sh
# Script to sweep relative x and z positions of primary and sec.
# coils, for simulation of dependence of link k on position.

FREQ=6.78e6
echo "# Link k calculation by fasthenry with possweep.sh">link.dat
# Initial run to include L1 and L2
echo "# Primry and sec L at large link distance:" >> link.dat
./fhgen $FREQ 1000.0 1000.0 > link.fhy
fasthenry link.fhy > /dev/null
echo "# L1: " $(MakeLcircuit Zc.mat 2>/dev/null | \
        grep "LZ_0 1" | cut -c19-31) >> link.dat
echo "# L2: " $(MakeLcircuit Zc.mat 2>/dev/null | \
        grep "LZ_1 3" | cut -c19-31) >> link.dat

echo "\n# x(mm) z(mm) k\n" >> link.dat

xpos=-40
zpos=0
while [ $xpos -le 80 ]
do
  (  while [ $zpos -le 140 ]
  do
    echo "Now calculating at x=" $xpos ", y=0, z=" $zpos
    ./fhgen $FREQ $xpos $zpos > link.fhy
    fasthenry link.fhy > /dev/null
    echo $xpos $zpos $(MakeLcircuit Zc.mat 2>/dev/null | \
        grep "KZ_1_0" | cut -c17-31) >> link.dat
    zpos=`expr $zpos + 4`
  done )
echo ' ' >> link.dat
xpos=`expr $xpos + 4`
done
```

B
PLL SIMULATION

Phase-locked loops are among the most difficult circuits to simulate because of the wide range of time constants. Modern loops typically use a digital phase-frequency detector, possibly with modifications to eliminate dead time [54, 3] and to shorten the time to lock [61]. The analog signals inside the phase detector change on a time scale of picoseconds, while the loop filter has time constants spanning several microseconds or even milliseconds. Because of the way in which traditional analog circuit simulators work, the entire simulation proceeds with a step size appropriated to the fastest signals, requiring an extraordinary number of steps for the whole simulation. There exist recent mixed-mode simulators, for example

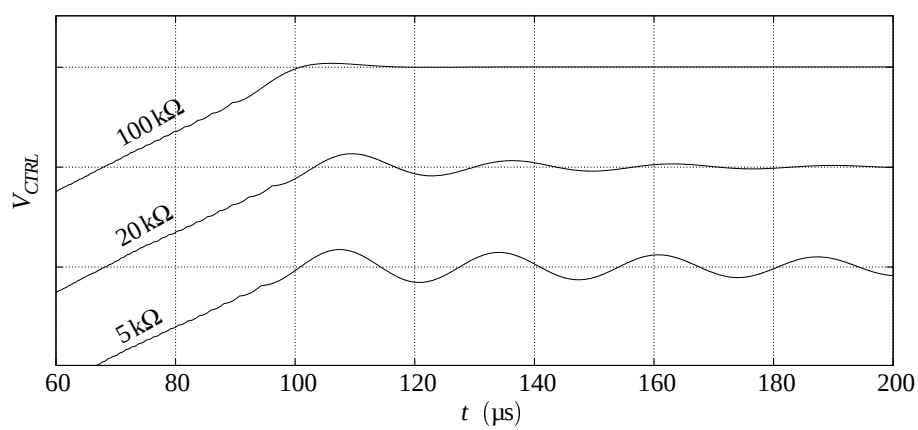


Figure B.1: *Simulation results for a PLL with a phase-frequency detector, and a standard loop filter as described by Gardner [38]. The pump current was 0.5 µA, the filter capacitor 50 pF, and the VCO gain 35Mrad/Vs. The simulation was run with three values of the filter resistor R2. The scales of the thee plots are offset from each other for clarity. The VCO control voltage is not calculated explicitly during the simulation, so the figure shows an internal filter node voltage instead. Because this is a low-pass filtered version of the control voltage, the cycle slip effects are not as visible as they could be.*

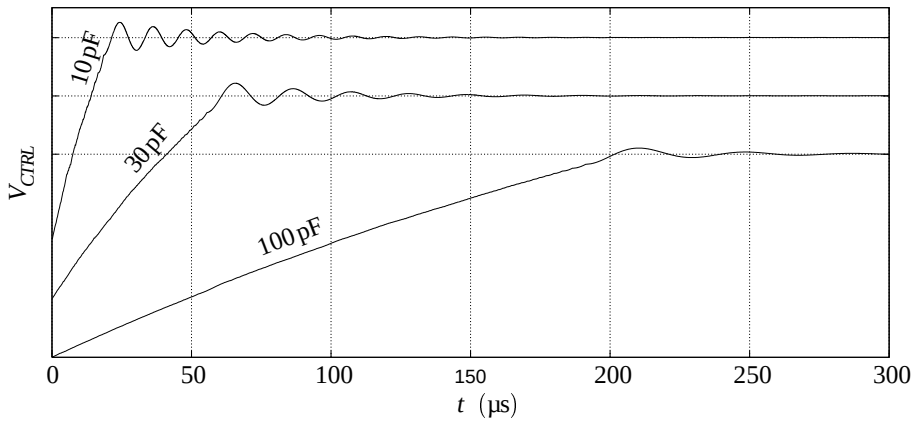


Figure B.2: *A similar set of results as in figure B.1, except that the filter resistor was fixed at 20 kΩ and the values of the filter capacitor were varied.*

based on VHDL-AMS circuit descriptions, which simplify the simulation task by abstracting complex blocks like the phase detector into a high-level macromodel. Experience with these tools does however indicate that they are still relatively slow and inaccurate.

Different simulator principles have been presented which address specifically the problem of simulating mixed-signal PLLs [7, 103, 62, 53]. The scheme used here is based on a method devised by Larsson in [60, 62]. The simulation program is based on an event prediction loop which predicts the next phase detector or clock event based on the current state. Internal analog signals such as the loop filter node voltages are described by floating-point variables and their time evolution is determined by a set of transfer functions. A sample program listing is given below for a loop consisting of a VCO, a simple PFD and a passive RC loop filter. Simulation results for a few typical runs are shown in figures B.1 and B.2. The time needed for each of these runs was 20 milliseconds, while a corresponding simulation in a circuit simulator like Spice or Spectre took a few hours on the same computer. The speed improvement factor was therefore on the order of 10^5.

This simulation program makes it possible to quickly obtain numbers for transient loop parameters which cannot be derived from linear loop theory [38, 39]. The linear theory is only valid while the loop is locked, for small variations in the input signal phase. The theory does not account for phenomena like cycle slip, and is unable to predict time-to-lock.

The program can relatively easily be extended to include nonlinear phase detectors [61] (often used to reduce the lock time), more complex passive or active loop filters, and noise effects to simulated jitter or phase noise [103].

```
/*-------------------------------------------------------------*/
/* pllsim.c                                                    */
/* Main program for event-driven PLL simulation                */
/* (C) 2000, Gunnar Gudnason                                   */
/*-------------------------------------------------------------*/
#include        <math.h>
```

```
#ifndef PI
#define PI      3.14159265358979323846264338327950288419716939937 5
#endif

#define REF 0x0001
#define VCO 0x0002

int triwave(int toggle);
int pfd(int toggle);

main()
    {
    int rlogic, vlogic; /* Logic state of the ref and vco sig.  */
    double r,v;         /* Phase of reference signal and VCO    */
                        /* Are truncated when they reach PI.    */
    double t,tstep;     /* Simulation time, and step            */
    double tv, tr;      /* Time to VCO toggle, and ref toggle   */
                        /* These time variables are not         */
                        /* truncated after each step.           */
    double I1,I1max;    /* Charge pump output current           */
    double x2,x3;       /* Voltages within the loop filter      */
    double a,b,c;       /* Variables used in event prediction   */

    const double R1=20000;      /* Loop filter component values */
    const double C2=10e-12;
    const double C3=1e-12;
    const double Ip=.5e-6;
    const double fr=2.*PI*5e6;  /* Reference input frequency    */
    const double K0=35e6;       /* VCO gain                     */
    const double minstep=1e-11; /* Minimum time step 10ps       */
    const double tdelta=1e-15;  /* 1 fs "delta" time step       */

    long i,j;
    int p;                      /* charge pump integer output   */

    r=v=0.;
    t=0.;
    x3=0.2;
    rlogic=vlogic=0;

    I1=Ip;
/*    I1max=2*Ip;*/        /* Factor 2 for modified triwave det.*/
    I1max=Ip;

    tr = t + (PI-r)/fr;
    a=(K0*I1max)/(2.*C2);
    b=K0*(x3 + I1max*R1);
    c=v-PI;
    tv = t - (b - sqrt(b*b - 4.*a*c))/(2.*a);

    for(i=0;t<300e-6;i++)
        {
        if(tr<tv)
            tstep = tr-t + tdelta;
        else
            {
            tstep=tv-t;
            if(tstep<minstep)tstep=minstep;
            }
        /* Forward equations update */
        t+=tstep;
```

```
        r += tstep*fr;
        v += K0*tstep*((x3+I1*R1) + (I1/(2.*C2))*tstep);
        x3+= (I1/C2)*tstep;

        /* Predict */
        if(r>=PI)
            {
            r-=PI;
            rlogic ^= 0x01;
            tr = t + (PI-r)/fr;
/*          p=triwave(REF);*/
            p=pfd(REF);
            I1=Ip*p;
            }
        else
            {
            if(v>=PI)
                {
                v-=PI;
                vlogic ^= 0x01;
/*              p=triwave(VCO);*/
                p=pfd(VCO);
                I1=Ip*p;
                }
            /* find lower bound on tv by using max. pump current */
            a=(K0*I1max)/(2.*C2);
            b=K0*(x3 + I1max*R1);
            c=v-PI;
            tv = t - (b - sqrt(b*b - 4.*a*c))/(2.*a);
            }
        if(!(i%10)) printf("%le %le\n",t*1e6,x3*1e3);
        }

    }
```

```
/*-----------------------------------------------------------------*/
/* pd.c                                                            */
/* Phase detectors for event-driven PLL simulation                 */
/* (C) 2000, Gunnar Gudnason                                       */
/*-----------------------------------------------------------------*/
#include <math.h>
#ifndef PI
#define PI      3.14159265358979323846264338327950288419716939937 5
#endif

#define REF 0x0001
#define VCO 0x0002

/* Modified triwave phase detector */
int triwave(int toggle)
    {
    int out;
    static int v=0,nv=1;
    static int r=0,r2=0,r3=0,r4=0,r5=0;
    static int u1=0,u2=0,u3=0,u4=0;

    if(toggle==REF)     /* Input signal event */
        {
        r^=0x0001;
        u1 = r^r2;
```

```
        }
    else                /* VCO event */
        {
        v^=0x0001;
        nv^=0x0001;
        if(v) /* Was rising VCO edge */
            {
            r2=r;
            r4=r3;
            }
        else /* falling edge */
            {
            r3=r2;
            r5=r4;
            }
        u1 = r^r2;
        u2 = r2^r3;
        u3 = r3^r4;
        u4 = r4^r5;
        }
    out = u1 - u2 - u3 + u4;
    return(out);
    }

/* Simple Phase-frequency detector based on two dff with        */
/* asynchronous reset, and one and gate.                        */
int pfd(int toggle)
    {
    static int r=0, v=0, u=0, d=0;

    if (toggle==REF)
        {
        r^=1;
        if(r==1)
            u=1;
        }
    else
        {
        v^=1;
        if(v==1)
            d=1;
        }
    if(u&d) /*reset */
        d=u=0;

    return(u-d);
    }
```

BIBLIOGRAPHY

[1] Asad A. Abidi. "Direct-conversion radio transceivers for digital communications". *IEEE J. Solid-State Circuits*, SC-30:1399–1410, December 1995.

[2] Pieter L. D. Abrie. *"Design of RF and microwave amplifiers and oscillators"*. Artech House, ISBN 0-89006-797-X, 1999.

[3] R. Ahola, J. Routama, S. Lindfors, and K. Halonen. "A phase detector with no dead zone and a very wide output voltage chargepump". In *Proc. ESSCIRC*, pages 155–158, 1998.

[4] T. Akin, K. Najafi, and R. M. Bradley. "A wireless implantable multichannel digital neural recording system for a micromachined sieve electrode". *IEEE J. Solid-State Circuits.*, SC-33:109–118, January 1998.

[5] P. J. Amantia et al. "A mixed-signal IC for a semi-implantable hearing aid". In *Proc. ISCAS*, June 1999.

[6] G. E. Loeb amd R. A. Peck. "Cuff electrodes for chronis stimulation and recording of peripheral nerve activity". *J. Neuroscience Methods*, 64:95–103, 1996.

[7] B. A. A. Antao, F. M. El-Turky, and R. H. Leonowich. "Mixed-mode simulation of phase-locked loops". In *Proc. Custom Integrated Circ. Conf.*, pages 8.4.1–8.4.4, 1993.

[8] P. Arcioni et al. "An innovative modelization of loss mechanism in silicon integrated inductors". *IEEE Trans. CAS II*, 46:1453–1460, December 1999.

[9] Jeffrey A. Von Arx and Khalil Najafi. "On-chip coils with integrated cores for remote inductive powering of integrated microsystems". In *The Ninth International Conference on Solid-State Sensors and Actuators*, June 1997.

[10] N. W. Ashcroft and N. D. Mermin. *"Solid state physics"*. Saunders College Publ., ISBN 0-03-049346-3, 1976.

[11] J. M. Audy. "3rd order curvature corrected bandgap cell". In *Proc. 38th Midwest Circuits and Systems Symposium*, pages 397–400, 1995.

[12] J. Bastos, M. Steyaert, B Graindourze, and W. Sansen. "Matching of MOS transistors with different layout styles". In *Proc. IEEE Conf. Microelectronics Test Structures*, pages 17–18, 1996.

[13] S. Benedetto, E. Biglieri, and V. Castellani. *"Digital transmission theory"*. Prentice Hall , ISBN 0-13-214313-5, 1987.

[14] Keith Billings. *"Switchmode power supply handbook"*. McGraw-Hill, ISBN 0-07-006719-8, 1999.

[15] David M. Binkley, James M. Rochelle, and Brian K. Swann. "A micropower CMOS direct-conversion VLF receiver chip for magnetic-field wireless applications". In *Proc. Custom Integrated Circ. Conf.*, pages 415–418, 1997.

[16] S. Bourret, M. Sawan, and R. Plamondon. "Programmable high-amplitude balanced stimulus current source for implantable microstimulators". In *Proc. IEEE EMBS Conf.*, pages 1938–1941, 1997.

[17] J. Bouvier et al. "A smart card CMOS circuit with magnetic power and communication interface". In *Proc. ISSCC.*, pages 296–297, 1997.

[18] J. S. Brugler. "Silicon transistor biasing for linear collector current temperature dependence". *IEEE J. Solid-State Circuits*, SC-2:57–58, June 1967.

[19] Erik Bruun. "CMOS current conveyors". In *Proc. ISCAS, pre-conference tutorial*, pages 632–641, 1994.

[20] Erik Bruun and Peter Shah. "Dynamic range of low-voltage cascode current mirrors". In *Proc. ISCAS*, pages 1328–1331, 1995.

[21] Mark Clements et al. "An implantable neuro-stimulator device for a retinal prosthesis". In *Proc. ISSCC*, pages 216–217, 1999.

[22] Mark Clements et al. "An implantable power and data receiver and neuro-stimulus chip for a retinal prosthesis system". In *Proc. ISCAS*, pages I–194, 1999.

[23] C. Cohen-Tannoudji, B. Liu, and F. Laloë. *"Quantum mechanics, Vol. 2"*. Wiley, ISBN 0-471-16435-6, 1977.

[24] Samuel D. Conte. *"Elementary numerical analysis"*. McGraw-Hill, 1980.

[25] Linear Technology Corporation. "LT1004 Micropower voltage reference data sheet". November 1998.

[26] National Semiconductor Corporation. "LM4041 precision micropower shunt voltage reference data sheet". January 2001.

[27] N. de N. Donaldson and T.A. Perkins. "Analysis of resonant coupled coils in the design of radio frequency transcutaneous links". *Medical and Biological Engineering and Computing*, 21:612–627, September 1983.

[28] C. J. Debono, F. Maloberti, and J. Micallef. "Low-voltage CMOS four-quadrant analogue multiplier for RF applications". *Electronics Letters*, 34:2285–2286, 1998.

[29] A. Deiss, D. Pfaff, and Q. Huang. "A 200MHz sub-mA RF front-end for wireless hearing aid applications". In *Proc. ESSCIRC*, 1999.

[30] A. Diaz-Sanchez, J. Ramirez-Angulo, E. Sanchez-Sinencio, and G. Han. "A CMOS four-quadrant current/transconductance multiplier". *Analog Integrated Circuits and Signal Processing*, 19:163–168, 1999.

[31] A. Djemouai, M. Sawan, and M. Slamani. "An efficient RF power transfer and bidirectional data transmission to implantable electronic devices". In *Proc. ISCAS*, pages II–259, 1999.

[32] J. F. Duque-Carillo. "Control of the common-mode component in CMOS continuous-time fully differential signal processing". *Analog Integrated Circuits and Signal Processing*, 4:131–140, 1993.

[33] Sayed-Amr El-Hamamsy. "Design of high-efficiency RF class-D power amplifier". *IEEE Trans. Power Electronics*, 9:297–308, May 1994.

[34] Christian C. Enz. *"The EKV model: A MOST model dedicated to low-current and low-voltage analogue circuit design and simulation"*. IEE Circuits and Systems Series 8. Peter Peregrinus Ltd., 1996.

[35] M. Felder and J. Granger. "Analysis of ground-bounce induced substrate noise coupling in a low-resistive bulk epitaxial process: design strategies to minimize noise effects on a mixed-signal chip". *J. Solid-State Circuits*, 46:1427–1436, 1999.

[36] Daniel P. Foty. *"MOSFET modeling with SPICE: principles and practice"*. Prentice-Hall , ISBN 0-13-227935-5, 1996.

[37] Douglas C. Galbraith, Mani Soma, and Robert L. White. "Wide-band efficient inductive transdermal power and data link with coupling insensitive gain". *IEEE Trans. Biomed. Eng.*, BME-34:265–275, April 1987.

[38] Floyd M. Gardner. "Charge-pump phase-lock loops". *IEEE Trans. Communications*, COM-28:1849–1858, November 1980.

[39] Floyd M. Gardner. "Phase accuracy of charge pump PLL's". *IEEE Trans. Communications*, COM-30:1362–2363, October 1982.

[40] Randall L. Geiger, Phillip E. Allen, and Noel R. Strader. *"VLSI design techniques for analog and digital integrated circuits"*. McGraw-Hill, ISBN 0-07-100728-8, 1990.

[41] Barrie Gilbert. "Current-mode circuits from a translinear viewpoint: A tutorial". In *Analogue IC design: the current-mode approach*. IEE Circuit and Systems Series 2, 1990.

[42] Barrie Gilbert. "Design practices for manufacturing robustness". In *Course material, Practical aspects in analog and mixed-mode design, EPFL (Lausanne)*, 1998.

[43] A. Giraldo, A. Paccagnella, and A. Minzoni. "Aspect-ratio calculation in n-channel MOSFETs with a gate-enclosed layout". *J. Solid-State Electronics*, 44:981–989, 2000.

[44] L. Glasser, A. C. Malamy, and C. W. Selvidge. "A magnetic power and communication interface for a CMOS integrated circuit". *IEEE J. Solid-State Circuits*, SC-24:1146–1149, August 1987.

[45] G. Gudnason. "Measurements and calculations of turbulent boundary layer evolution over smooth and wavy surfaces in the CWT-1". Technical report, Von Karman Institute for Fluid Dynamics, 1990.

[46] G. Gudnason. "An implantable multichannel neural stimulator for functional electrical stimulation". Master's thesis, The Technical University of Denmark, 1997.

[47] G. Gudnason, E. Bruun, and M. Haugland. "A chip for an implantable neural stimulator". *Analog Integrated Circuits and Signal Processing*, 22:81–89, 2000.

[48] M. Gunawan et al. "A curvature-corrected low-voltage bandgap reference". *IEEE J. Solid-State Circuits*, SC-28:667–670, June 1993.

[49] Richard Haberman. *"Elementary applied partial differential equations"*. Prentice-Hall, ISBN 0-13-252891-6, 1987.

[50] Q. Huang and M. Oberle. "A 0.5 mW passive telemetry IC for biomedical applications". *IEEE J. Solid-State Circuits.*, SC-33:937–946, July 1998.

[51] M. Ilić, D. Vasiljević, and D. B. Popović. "A programmable electronic system for FES systems". *IEEE Trans. Rehab. Eng.*, 4:234–239, December 1994.

[52] J. D. Jackson. *"Classical electrodynamics, 2nd ed."*. Wiley , ISBN 0-471-43132-X, 1975.

[53] Norman K. James. "Cycle-domain simulator for phase-locked loops". In *Proc. Southwest Symp. on Mixed-Signal Design*, pages 77–82, 2000.

[54] Henrik O. Johansson. "A simple precharged CMOS phase frequency detector". *IEEE J. Solid-State Circuits*, SC-33:295–299, February 1998.

[55] David A. Johns and Ken Martin. *"Analog integrated circuit design"*. Wiley, ISBN 0-471-14448-7, 1997.

[56] M. Kamon, M.J. Tsuk, and J.K. White. "FASTHENRY: A multipole-accelerated 3-D inductance extraction program". *IEEE Trans. Microwave Theory and Techniques*, 31:1750–1758, Sept 1994.

[57] Mattan Kamon. "Efficient techniques for inductance extraction of complex 3-D geometries". Master's thesis, Massachusetts Institute of Technology, 1994.

[58] S. Kolnsberg et al. "Microelectronic components for a retina implant system". In *Proc. ESSCIRC*, 1999.

[59] Kenneth R. Laker and Willy M. C. Sansen. *"Design of analog integrated circuits and systems"*. McGraw-Hill, ISBN 0-07-113-458-1, 1994.

[60] Patrik Larsson. "Fast and accurate event driven simulation of partly analog phase-locked loops". In *Proc. ISCAS*, pages 1756–1759, 1995.

[61] Patrik Larsson. "Reduced pull-in time of phase-locked loops using a simple nonlinear phase detector". *IEE Proceedings on Communications*, 142:221–226, August 1995.

[62] Patrik Larsson. "A simulator core for charge-pump PLLs". *IEEE Trans. CAS II*, CAS-45:1323–1326, September 1998.

[63] Thomas H. Lee. *"The design of CMOS radio-frequency integrated circuits"*. Cambridge University Press , ISBN 0-521-63922-0, 1998.

[64] T. Lehmann. "Using current-steering logic in mixed analogue-digital circuits". *Electronics Letters*, 34:1660–1661, August 1998.

[65] S. Liu, C. Chang, and Y. Hwang. "New CMOS four-quadrant multiplier with squarer circuit". *Analog Integrated Circuits and Signal Processing*, 9:257–263, 1996.

[66] G. E. Loeb et al. "Injectable microstimulator for functional electrical stimulation". *Medical and Biological Engineering and Computing*, 29:NS13–NS19, November 1991.

[67] S. S. Mohan, M. Hershenson, S. P. Boyd, and T. H. Lee. "Simple accurate expressions for planar spiral inductances". *IEEE J. Solid-State Circuits*, SC-34:1419–1424, October 1999.

[68] M. Nardin and K. Najafi. "A Multichannel neuromuscular microstimulator with bi-directional telemetry". In *Proc. Conf. Solid-State Sensors and Actuators*, pages 59–62, June 1995.

[69] Bram Nauta. *"Analog CMOS filters for very high frequencies"*. Kluwer Academic Publishers, ISBN 0792392728, 1993.

[70] G. Nebel et al. “ASK 10% demodulator for contactless smart card IC”. In *Proc. ESSCIRC*, pages 230–233, 1999.

[71] Hiok-Tiaq Ng and David J. Allstot. “CMOS current steering logic for low-voltage mixed-signal integrated circuits”. *IEEE Trans. VLSI*, 5:301–308, September 1997.

[72] Jannik H. Nielsen and Torsten Lehmann. “An implantable CMOS amplifier for nerve signals”. In *Proc. ICECS*, pages 1183–1186, September 2001.

[73] A. M. Niknejad. R.G. Meyer Research Group, UC Berkeley, `http://formosa.eecs.berkeley.edu/~niknejad/asitic.html`.

[74] A. M. Niknejad and R. M. Mayer. “Analysis of eddy-current losses over conductive substrates with applications to monolithic inductors and transformers”. *IEEE Trans. Microwave Theory and Techniques*, 49:166–176, January 2001.

[75] Paul O’Leary. *“Practical aspects of mixed analogue and digital design”*. IEE Circuits and Systems Series 3. Peter Peregrinus Ltd., ISBN 0-86341-259-9, 1991.

[76] O. Oliaei and P. Loumeau. “Four-quadrant class AB CMOS current multiplier”. *Electronics Letters*, 32:2327–2329, 1996.

[77] K. Papathanasiou and T. Lehmann. “An implantable CMOS signal conditioning system for recording nerve signals with cuff electrodes”. In *Proc. ISCAS*, May 2000.

[78] Bob Pease. “The design of band-gap reference circuits: trials and tribulations”. In *IEEE Proc. Bipolar Circuits and Technology Meeting*, 1990.

[79] J. Popović, A. Pavasović, and D. Vasiljević. “Low-power CMOS current-conveyor relaxation oscillators”. *IEEE Trans. Ultrasonics, Ferroelectronics and Freq. Control*, 44:895–901, July 1997.

[80] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *“Numerical recipes in C”*. Cambridge University Press, ISBN 0-521-43108-5, 1992.

[81] Maxim Integrated Products. “MAX6381-MAX6390 single/dual low-voltage, low-power µP reset circuits data sheet”. September 2000.

[82] S. Ramo, J. R. Whinnery, and T. Van Duzer. *“Fields and waves in communication electronics, 3rd ed.”*. Wiley , ISBN 0-471-30578-2, 1994.

[83] Behzad Razavi. *“RF microelectronics”*. Prentice-Hall , ISBN 0-13-887571-5, 1998.

[84] Research Laboratory of Electronics, Massachusetts Institute of Technology, `ftp://rle-vlsi.mit.edu/pub/fasthenry/`.

[85] G. Rincon-Mora and P. E. Allen. “A 1.1-V current-mode and piecewise-linear curvature-corrected bandgap reference”. *IEEE J. Solid-State Circuits*, SC-33:1551 –1554, October 1998.

[86] G. A. Rincon-Mora and P. E. Allen. “A low-voltage, low quiescent current, low drop-out regulator”. *IEEE J. Solid-State Circuits*, SC-33:36–44, January 1998.

[87] L. S. Robblee and T. L. Rose. “Electrochemical guidelines for selection of protocols and electrode materials for neural stimulation”. In *Neural Prostheses: Fundamental Studies*, chapter 2, pages 25–66. Prentice Hall, 1989.

[88] T. L. Rose and L. S. Robblee. “Electrical stimulation with Pt electrodes. VIII. Electrochemically safe charge injection limits with 0.2ms pulses”. *IEEE Trans. Biomedical Engineering*, BME-37:1118–1120, November 1990.

[89] R. T. L. Saez et al. “CMOS current steering logic: towards a mature technique for mixed-mode applications”. In *Proc. CICC*, pages 349–352, 1997.

[90] R. T. L. Saez et al. “Design guidelines for CMOS current steering logic”. In *Proc. ISCAS*, pages 1872–1875, 1997.

[91] K. Salmi et al. “4V, 5mA low dropout regulator using series-pass n-channel MOSFET”. *Electronics Letters*, 35:1214–1215, July 1999.

[92] M. Sawan, S. Robin, B. Provost, Y. Eid, and K. Arabi. “A wireless implantable electrical stimulator based on two FPGAs”. In *Proc. ICECS*, October 1996.

[93] Mohammed Sawan et al. “Computerized transcutaneous control of a multichannel implantable urinary prosthesis”. *IEEE Trans. Biomedical Engineering*, BME-39:600–609, June 1992.

[94] Mohammed Sawan et al. “Stimulator design and subsequent stimulation parameter optimization for controlling micturition and reducing urethral resistance”. *IEEE Trans. Rehabilitation Engineering*, 4:39–46, March 1996.

[95] Evert Seevinck. “Analogue interface circuits for VLSI”. In *Analogue IC design: the current-mode approach*. IEE Circuit and Systems Series 2, 1990.

[96] National Semiconductor. “LP3470 Tiny Power On Reset Circuit Data Sheet”. September 2000.

[97] H. J. Shin et al. “Low drop-out on-chip voltage regulator for low-power circuits”. In *Proc. IEEE Symp. Low Power Electronics*, pages 76–77, 1994.

[98] Brian Smith et al. “An externally powered, multichannel, implantable stimulator-telemeter for control of paralyzed muscle”. *IEEE Trans. Biomedical Engineering*, BME-45:463–475, 1998.

[99] Jack Smith. “*Modern communication circuits*”. McGraw-Hill, ISBN 0-07-066544-3, 1986.

[100] R. St-Amand, M. Sawan, and Y. Savaria. “Design and optimization of a low DC offset CMOS current source dedicated to implantable miniaturized stimulators”. In *Analog Integrated Circuits and Signal Processing*, pages 47–61, 1996.

[101] A. Van Staveren, J. Van Velzen, C. J. M. Verhoeven, and A. H. M. Van Roermund. “An integratable second-order compensated bandgap reference for 1V supply”. *Analog Integrated Circuits and Signal Processing*, pages 69–81, 1995.

[102] Z. Tang et al. “Data transmission form an implantable biotelemeter by load-shift keying using a circuit configuration modulator”. *IEEE Trans. Biomedical Engineering*, BME-42:524, May 1995.

[103] W. E. Thain and J. A. Connelly. “Simulating phase noise in phase-locked loops with a circuit simulator”. In *Proc. ISCAS*, pages 1760–1763, 1995.

[104] M. Thomsen. “*Characterization and optimization of whole nerve recording cuff electrodes*”. PhD thesis, University of Aalborg, Denmark, 1998.

[105] M. Toki. “Low DC-driving-voltage crystal oscillator”. In *Proc. Joint meeting EFTF–IEEE IFCS*, pages 412–415, 1999.

[106] Philip R. Troyk and Martin A. K. Schwan. “Class E driver for transcutaneous power and data link for implanted electronic devices”. *Medical and Biological Engineering and Computing*, 30:69–75, 1992.

[107] Philip R. Troyk and Martin A. K. Schwan. “Closed-loop class E transcutaneous power and data link for microimplants”. *IEEE Trans. Biomedical Engineering*, BME-39:589, June 1992.

[108] Y. P. Tsividis. “Accurate analysis of temperature effects in I_C-V_{BE} characteristics with applications to bandgap reference sources”. *IEEE J. Solid-State Circuits*, SC-15:1076–1084, December 1980.

[109] Yannis Tsividis. *“Operation and modeling of the MOS transistor, 2nd ed”*. McGraw-Hill , ISBN 0-07-065523-5, 1999.

[110] G. Tzanateas, C. A. T. Salma, and Y. P. Tsividis. “A CMOS bandgap voltage reference”. *IEEE J. Solid-State Circuits*, SC-14:655–657, June 1979.

[111] UC Berkeley Device Group, BSIM Homepage, `http://www-device.EECS.Berkeley.EDU/~bsim3/`.

[112] J. A. T. M. van den Homberg. “A universal $0.03\,\mathrm{mm}^2$ one-pin crystal oscillator in CMOS”. *IEEE J. Solid-State Circuits*, SC-34:956–961, July 1999.

[113] A. C. van der Woerd et al. “Low-power current-mode 0.9-V voltage regulator”. *IEEE J. Solid-State Circuits*, SC-29:1138–1141, September 1994.

[114] C. Veraart, W. Grill, and T. Mortimer. “Selective control of muscle activation with a multipolar cuff electrode”. *IEEE Trans. BME*, BME-40:640–653, 1993.

[115] E. A. Vittoz, M. G. R. Degrauwe, and S. Bitz. “High-performance crystal oscillator circuits: theory and application”. *IEEE J. Solid-State Circuits*, SC-23:774–783, June 1988.

[116] Eric A. Vittoz. “Micropower Techniques”. In *Design of analog-digital VLSI circuits for telecommunications and signal processing*, chapter 3. Prentice Hall, ISBN 0-13-203639-8, 1994.

[117] R. K. Wangsness. *“Electromagnetic fields, 2nd ed.”*. Wiley , ISBN 0-471-81186-6, 1986.

[118] Neil H. E. Weste and Kamran Eshragian. *“Principles of CMOS VLSI design, 2nd ed”*. Addison-Wesley, ISBN 0-201-53376-6, 1994.

[119] R. J. Widlar. “New developments in IC voltage regulators”. *IEEE J. Solid-State Circuits*, SC-6:2–7, February 1971.

[120] R. J. Widlar. “Low voltage techniques”. *IEEE J. Solid-State Circuits*, SC-13:838–846, December 1978.

[121] K. A. Witzke and C. Leung. “A comparison of some error detecting CRC code standards”. *IEEE Trans. Communications*, COM-33:996–998, September 1985.

[122] B. Ziaie, M. D. Nardin, A. R. Coglan, and K. Najafi. “A single-channel implantable microstimulator for functional neuromuscular stimulation”. *IEEE Trans. Biomed. Eng.*, BME-44:909–919, October 1997.

[123] B. Ziaie, S. C. Rose, M. K. Nardin, and K. Najafi. “A self-oscillating detuning-sensitive class E transmitter for implantable microsystems”. *IEEE Trans. Biomedical Engineering*, BME-48:397–400, March 2001.

PUBLICATIONS

The following journal and conference articles were presented during the course of this study.

[1] Gunnar Gudnason, Erik Bruun, and Morten Haugland. "A chip for an implantable stimulator". *Proceedings of the 16th Norchip Conference*, pp. 308–315, 1998.

[2] Gunnar Gudnason, Erik Bruun, and Morten Haugland. "An implantable mixed analog/digital neural stimulator circuit". *Proc. ISCAS*, pp. 375–378, 1999.

[3] Gunnar Gudnason, Erik Bruun, and Morten Haugland. "A chip for an implantable neural stimulator". *Analog Integrated Circuits and Signal Processing*, **22**:81–89, 2000.

[4] Gunnar Gudnason. "A low-power ASK demodulator for inductively coupled implantable electronics". *Proc. ESSCIRC*, pp. 404–407, 2000.

[5] Gunnar Gudnason, Jannik H. Nielsen, Erik Bruun, and Morten Haugland. "A distributed transducer system for functional electrical stimulation". *Proc. ICECS*, pp. 397–400, 2001.

The articles are reproduced on the following pages, by permission of their respective copyright holders.

A CHIP FOR AN IMPLANTABLE NEURAL STIMULATOR

Gunnar Gudnason*, Erik Bruun* and Morten Haugland**

*Department of Information Technology, Building 344, DTU, DK–2800 Lyngby,
Tel.: +45 45 25 39 06, Fax: +45 45 88 01 17, E-mail: gg@it.dtu.dk, bruun@it.dtu.dk

**Center for Sensory-Motor Interaction, Aalborg University, Fr. Bajersvej 7D, DK–9220 Aalborg, Tel.: +45 96 35 88 24, Fax: +45 98 15 40 08, E-mail: mh@miba.auc.dk

Abstract

This paper describes a chip for a multichannel neural stimulator for functional electrical stimulation. The chip performs all the signal processing required in an implanted neural stimulator. The power and signal transmission to the stimulator is carried out via an inductive link. From the signals transmitted to the stimulator, the chip is able to generate charge-balanced current pulses with a controllable length and amplitude for stimulation of nerve fibres. The chip has 4 output channels so that it can be employed in a cuff electrode with multiple connections to a nerve. The purpose of the functional electrical stimulation is to restore various bodily functions (e.g. motor functions) in patients who have lost them due to injury or disease.

INTRODUCTION

Functional electrical stimulation is the activation of physical functions through electrical stimulation of nerve tissue. It can be applied for instance to victims of spinal chord injuries who have lost control over part of their body (paralysis) or to patients suffering from foot drop. By stimulating the nerves controlling the muscles, some degree of muscular control is regained. Several implantable solutions to this problem have been described in the literature, e.g. [1, 2], each with their distinctive features. For the stimulator described here, an important consideration is the physical size and the option of having several output channels. The stimulator is an implantable unit comprising electrodes to contact the nerves, a signal processing chip to generate the stimulation pulses, and a coil (and some few other discrete devices) to provide an inductive data and power link to the stimulator. To keep the size small it is not only necessary to minimize the number of discrete components but also to pay attention to the size of the signal processing

Reprinted with permission from the proceedings of the 1998 NORCHIP conference.

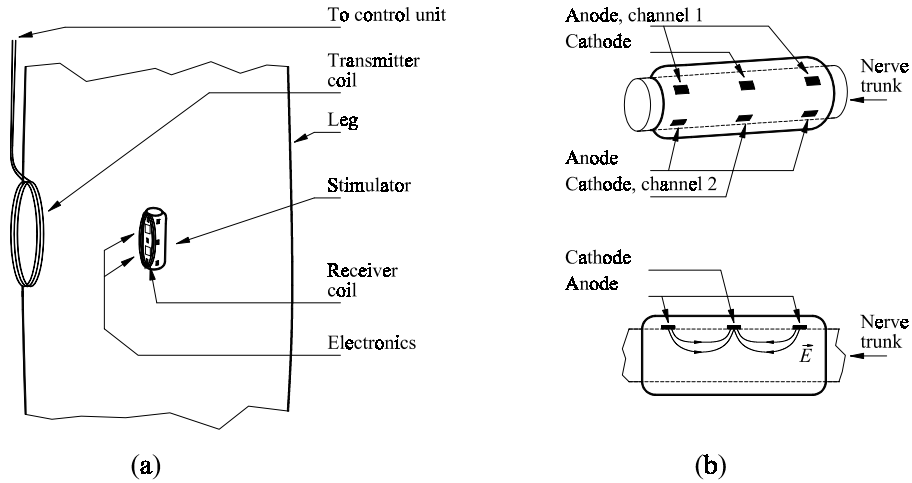


Figure 1: Placement of implanted stimulator and transmitter coil on the lower leg (a). Cuff electrode (perspective view and electric field during stimulation) (b).

chip. In addition to the implanted stimulator, the system comprises a control unit and a transmitter coil placed on the surface of the skin as shown in fig. 1(a). The implanted stimulator is a small unit placed around a nerve fibre with so-called cuff electrodes to connect the stimulator outputs to the nerve tissue. Fig. 1(b) shows the cuff electrode.

SYSTEM SPECIFICATION

Some of the important system requirements are the following: The system must be powered via the inductive link. The stimulation pulses are current pulses which must be programmable in amplitude and duration and each stimulation pulse must be followed by a pulse in the reverse direction ensuring that no charge build-up takes place in the nerve tissue. The maximum amplitude is $2mA$ and the maximum pulse duration is $255\mu s$. The nerve tissue can be expected to exhibit an ohmic resistance of up to $5k\Omega$, implying that the stimulator must be able to generate stimulation pulses of up to $10V$ amplitude. It must be possible to direct the stimulation pulse to a selected electrode among a total number of at least 4 electrodes.

To meet these requirements a low power signal processing chip is required with a unidirectional transmission protocol to program the stimulator pulses and a carrier and modulation scheme which ensures a sufficient power supply.

Fig. 2 shows the stimulator components. In addition to the chip and the coil, a tuning capacitor, a rectifier with a filter capacitor and a capacitor for the charge balancing of the stimulation pulses are required. Also, an external zener diode, D_Z, provides a shunt regulation and overvoltage protection for the chip.

TRANSMISSION PROTOCOL

For the power and data transfer to the stimulator, an inductive link with a carrier frequency of $5MHz$ has been chosen. This frequency is chosen as a compromise between

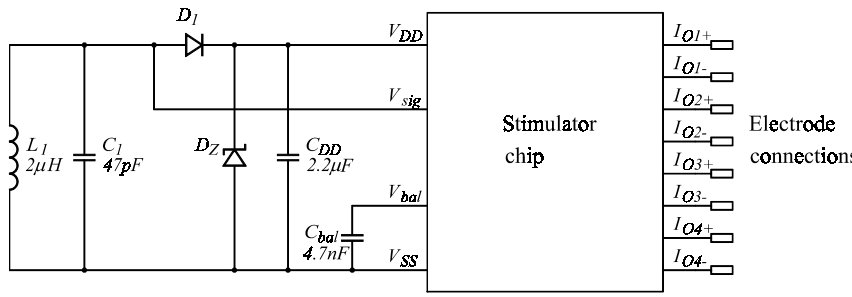


Figure 2: The stimulator components

inductive coupling (best at high frequencies), and small absorption of tissue and simple electronics (best at low frequencies). The carrier signal is rectified and filtered by the external diode D_1 and the capacitor C_{DD} shown in fig. 2. The supply voltage V_{DD} has been selected to $12V$ in order to be able to generate $2mA$ current pulses into a $5k\Omega$ load. The supply voltage is controlled by the external zener diode D_Z. The data signal is encoded by a pulse amplitude modulation of the carrier with on-off keying. The data rate has been selected to $100kbits/s$. This modulation has been selected because of its simplicity and robustness against imperfections in the transmission but it has the disadvantage that no clock or power is transmitted during the off keying. Hence, the clock recovery circuit must be able to maintain a stable clock frequency during the off keying intervals. For the control of pulse length and amplitude 8-bit words are chosen. Also, to select the output channel for the stimulation, an 8-bit word is chosen. This means that the same transmission protocol can be employed for stimulators with up to 255 output channels. The complete command word to the stimulator is as shown in fig. 3. In addition to the channel select, amplitude and duration control there are two start bits and a cyclic redundancy check word (CRC) for error detection.

THE CHIP BLOCK DIAGRAM

A simplified block diagram of the chip is shown in fig. 4. The chip comprises a number of analog circuits functions in addition to a digital control circuit and a digital to analog converter for the pulse amplitude control.

Voltage regulator

The voltage regulator contains a bandgap reference which is used to control a series regulator for a $3.3V$ supply to the digital blocks in the chip. The main reason for introducing the series regulator is to save power. The power consumption of CMOS logic increases with the square of the supply voltage. At $3.3V$ the CMOS logic is expected to draw

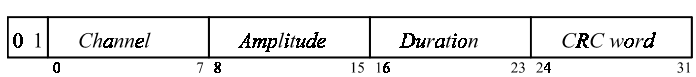


Figure 3: The command word

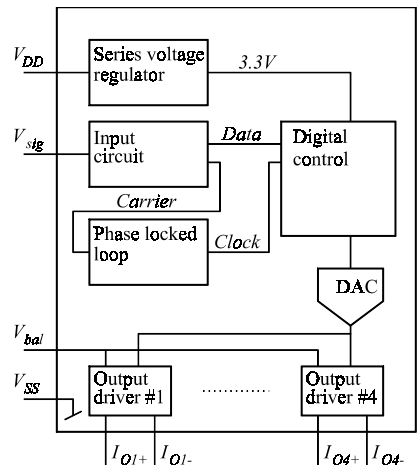


Figure 4: Chip block diagram

about $80\mu A$ leading to a power consumption of about $0.26mW$. With a $12V$ power supply the consumption would be about $3.4mW$, so the series regulator provides a significant power saving even though about $0.7mW$ is dissipated in the series pass transistor of the regulator. The series regulator is fairly conventional.

Input circuit and clock recovery

The input circuit serves the purpose of extracting the carrier for the generation of the system clock and detecting the modulation for retrieving the data transmitted to the stimulator. Basically, the clock is derived directly from the input signal by taking the signal through a clipping source follower stage, followed by a couple of inverters. The clock derived in this way is of course only detected when the amplitude of the input signal is high, corresponding to a logic 1. When the input is low, no clock signal can be detected since the modulation scheme is a simple on/off keying. For the purpose of providing a continuous clock signal to the chip a phase locked loop is employed. The loop locks to the carrier frequency during logic 1 transmissions. During logic 0 transmissions the VCO of the PLL holds the frequency. This is achieved by using a carrier detect signal

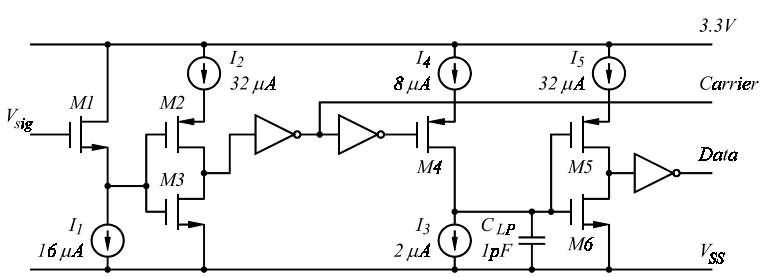


Figure 5: The data input stage

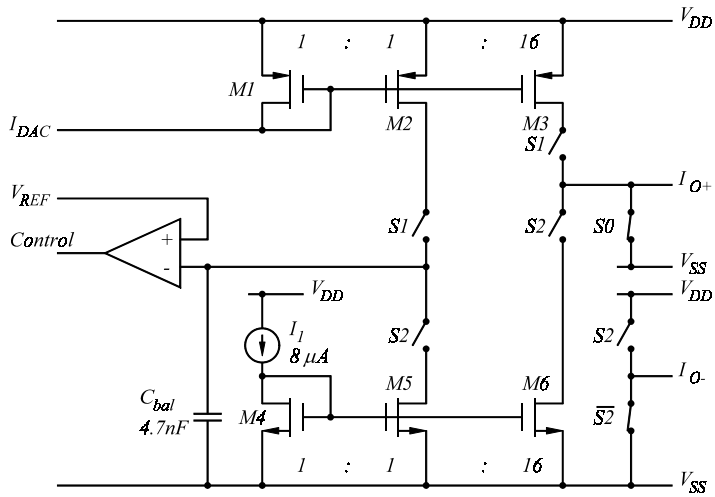


Figure 6: The output driver stage

to enable the PLL. Only when a carrier is detected will the PLL try to lock onto the input frequency. The output from the PLL is divided by 5 to provide a system clock of $1MHz$.

The envelope detection is done by means of the charge pump circuit shown in fig. 5. The capacitor C_{LP} is continuously discharged by the constant current source I_3. When the carrier is present (i.e. a logic 1 is transmitted), C_{LP} is charged through $M4$ and the current source I_4. The capacitor voltage is at equilibrium when the carrier duty cycle is equal to I_3/I_4 or 1/4. When the duty cycle is higher, as when the input signal is near maximum amplitude, the capacitor voltage tends to $3.3V$ (the digital supply voltage). Conversely, when the duty cycle is lower or even zero, the capacitor voltage tends to V_{SS}. A ripple of about $\Delta V = I_3/(2C_{LP}f_{carrier})$ will be present on the capacitor when the voltage is high due to the discharge when the clock is low. By selecting proper current levels ($I_3 = 2\mu A$) and capacitor size ($1pF$) the ripple can be kept very small and it is filtered out by the current limited inverter $M5 - M6$ and the following standard digital inverter.

Amplitude control and output driver

The pulse amplitude is controlled by an 8-bit command as shown in fig. 3. This means that it can be selected to one of 255 current levels ranging from $0 - 2mA$. The length of the pulse is also selected by an 8-bit command to be $1 - 255\mu s$. Each pulse is followed by a charge balancing pulse with an amplitude of $128\mu A$ and an appropriate length to achieve the charge balance.

A very simple way to achieve charge balance would be to use an ac-coupling (series capacitor) to the contacting electrodes. However, this would require an external capacitor for each of the output channels which is undesirable because of the space constraints of the stimulator. An alternative is to calculate the charge in the stimulation pulse and derive the pulse length of the charge balancing pulse from this calculation. This can be

done in the digital domain from the information received about the pulse amplitude and length, but the calculation would require a good deal of (area consuming) logic and the charge balancing pulse would be quantized in minimum steps of $128\mu A \times 1\mu s$. As an alternative, an analog calculation described below has been selected.

A simplified schematic of an output driver stage is shown in fig. 6. The switches $S1$ are 'on' when the output driver is selected and a pulse is activated. The switch $S0$ are 'on' when the output stage is idle. The switches $S2$ are 'on' when a charge balancing pulse is activated. Basically, the pulse is generated by the current mirror $M1, M3$. The input current is derived from a digital to analog converter with an output current range of $0-128\mu A$. The current source I_1 is used for the charge balancing pulse. During the output pulse, $M2$ charges the capacitor C_{bal} with a current equal to the output current divided by 16. During the discharge pulse, the capacitor C_{bal} is discharged with the discharge current divided by 16, i.e. $8\mu A$. Thus, charge balance is obtained when the capacitor voltage is returned (by I_{D5}) to its initial, precharged value detected by a comparator. The comparator output is used by the control logic to terminate the switch control signal S_2. In the actual implementation, the current mirrors $M1 - M3$ and $M4 - M6$ are implemented as cascode current mirrors to improve the accuracy. The charge balancing also depends on the matching in the output current mirrors. The maximum charge delivered by an output pulse is $Q_{max} = t_{max} \times I_{max} = 255\mu s \times 2mA \simeq 0.5\mu C$. With a mismatch on the order of 5% a charge error of about $25nC$ can be expected. This can be compared to the generally accepted levels of charge accumulation of $1 - 3\mu C/mm^2$ reported in [4] and [5]. The charge balancing capacitor C_{bal} has a value of $4.7nF$ which is certainly not integratable. However, a single external capacitor can be used for all the output channels.

The digital to analog converter must provide a digitally controllable current output in the range $0 - 128\mu A$. It is implemented as an array of equally sized current sources controlled by the digital command word.

Digital control

The digital control logic comprises all the logic needed to decode the command word shown in fig. 3 to the appropriate inputs to the analog to digital converter, the output channel decoder, and the timing of the output pulse. Also, the digital control performs the CRC-8 check of the command word. This reduces the risk of issuing false stimulation pulses. No error correction has been implemented since the application would normally allow a simple retriggering of the stimulation pulses if a pulse was missing due to a transmission error.

EXPERIMENTAL RESULTS

Prototype chips have been fabricates in MIETEC's $2\mu m$ CMOS technology. The process has been selected primarily because of its capability to handle the rather large supply voltage. The total chip size is $3mm \times 4mm$. For evaluation purposes the chip has been encapsulated in a standard dual-in-line package. A chip photo is shown in fig. 7. For devices to be used for implantation an alternative encapsulation and mounting together with the discrete devices must be employed.

The chip has been tested with respect to some critical parameters and with respect to functionality. All the circuit blocks have been found to work satisfactorily. The carrier

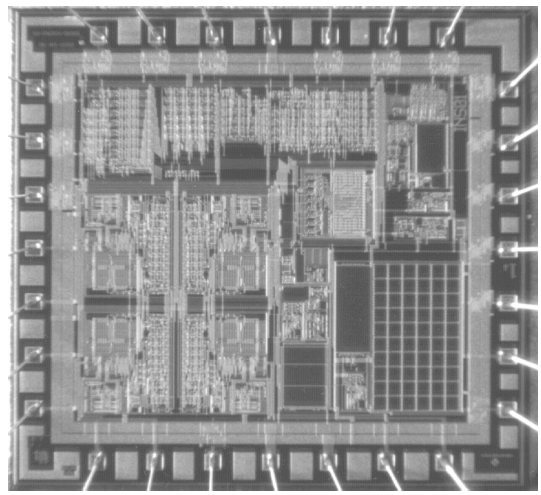


Figure 7: Chip photo of the stimulator chip

recovery and envelope detection work as predicted. The PLL is able to lock on the carrier and hold its frequency when the carrier is off with a decay of less than $0.1\%/s$. The digital supply voltage has been measured to be within a tolerance of 10%. The total supply current is about $600\mu A$ which is close to the simulated value. The charge balancing circuit shows a mismatch of 5–10% of the stimulation charge, slightly more than expected, but still acceptable.

Functionally, the chip works as expected. Fig. 8 shows the waveform of the output current when the output has been programmed to a duration of $255\mu s$ and an amplitude of $0.6mA$. The output has been measured as the voltage across a $10k\Omega$ resistor. The charge balancing pulse following the stimulation pulse is clearly seen.

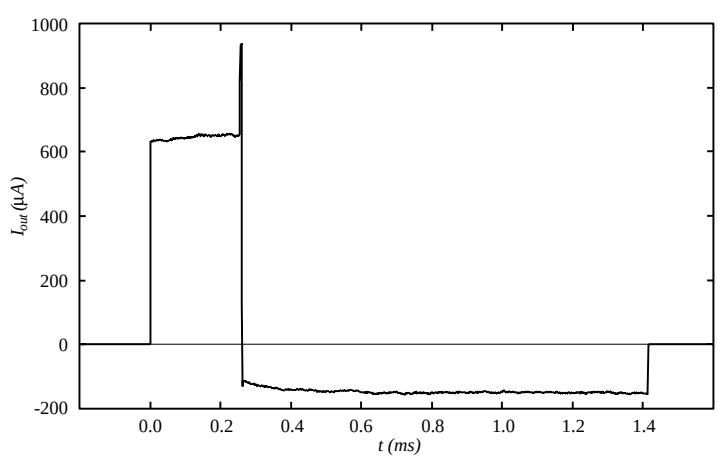


Figure 8: Measured output current pulse from the stimulator

CONCLUSION

A prototype chip for a neural stimulator has been designed and fabricated. The chip comprises all the signal processing functions required for the application. A novel method for charge balancing of the stimulation pulses has been developed. Initial measurements show full functionality of the chip and specifications well within the tolerances expected. The next step will be to employ the chip in experiments with functional electrical stimulation of rabbits in order validate that the performance specification of the chip are adequate for the planned applications.

Acknowledgement

The Siemens Foundation is acknowledged for providing financial support for the fabrication of the chip prototypes.

References

[1] M. Sawan, S. Robin, B. Provost, Y. Eid, and K. Arabi, "A Wireless Implantable Electrical Stimulator Based on Two FPGAs," *Proc. Third IEEE International Conf. on Electronics, Circuits and Systems*, pp. 1092-95, Rodos, Greece, October 1996.

[2] B. Smith, P. H. Peckham, M. W. Keith, and D. D. Roscoe "An Externally Powered, Multichannel, Implantable Stimulator for Versatile Control of Paralyzed Muscle," *IEEE Trans. Biomed. Eng.*, Vol. BME-34, pp. 499-598, July 1987.

[3] Gunnar Gudnason, *An Implantable Multichannel Neural Stimulator for Functional Electrical Stimulation,* M. Sc. thesis, Department of Information Technology, Technical University of Denmark, 1997.

[4] N. de N. Donaldson and P. E. K. Donaldson, "When are Actively Balanced Biphasic ('Lilly') Stimulating Pulses Necessary in a Neurological Prosthesis? " *Medical and Biological Engineering and Computing*, Vol. 24, pp. 41-56, January 1996.

[5] T. L. Rose and L. S. Robblee, "Electrical Stimulation with Pt Electrodes. VIII. Electrochemically Safe Charge Injection Limits with 0.2ms Pulses," *IEEE Trans. Biomed. Eng.*, Vol. BME-37, pp. 1118-1120, November 1990.

AN IMPLANTABLE MIXED ANALOG/DIGITAL NEURAL STIMULATOR CIRCUIT

Gunnar Gudnason[1] *Erik Bruun*[1] *Morten Haugland*[2]

[1]Technical University of Denmark, Dept. of Information Technology, Bldg. 344, DK–2800 Lyngby, Denmark
[2]Center for Sensory-Motor Interaction, Aalborg University, Fr. Bajersvej 7D, DK–9220 Aalborg, Denmark

ABSTRACT

This paper describes a chip for a multichannel neural stimulator for functional electrical stimulation. The chip performs all the signal processing required in an implanted neural stimulator. The power and signal transmission to the stimulator is carried out via an inductive link. From the signals transmitted to the stimulator, the chip is able to generate charge-balanced current pulses with a controllable length and amplitude for stimulation of nerve fibres. The chip has 4 output channels so that it can be employed in a cuff electrode with multiple connections to a nerve. The purpose of the functional electrical stimulation is to restore various bodily functions (e.g. motor functions) in patients who have lost them due to injury or disease.

1. INTRODUCTION

Functional electrical stimulation is the activation of physical functions through electrical stimulation of nerve tissue. It can be applied for instance to victims of spinal chord injuries who have lost control over part of their body (paralysis) or to patients suffering from foot drop. By stimulating the nerves controlling the muscles, some degree of muscular control is regained. Several implantable solutions to this problem have been described in the literature, e.g. [1, 2], each with their distinctive features. For the stimulator described here, an important consideration is the physical size and the option of having several output channels. The stimulator is an implantable unit comprising electrodes to contact the nerves, a signal processing chip to generate the stimulation pulses, and a coil

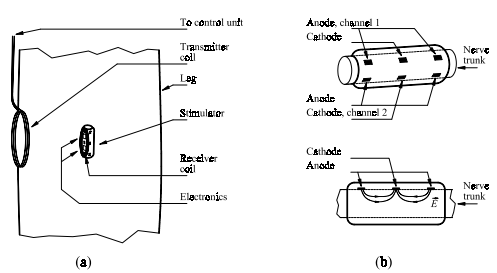


Figure 1. Placement of implanted stimulator and transmitter coil on the lower leg (a). Cuff electrode (perspective view and electric field during stimulation) (b).

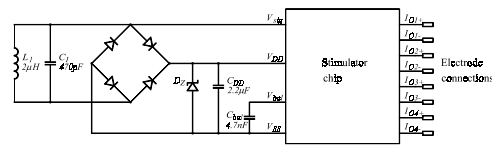


Figure 2. The stimulator components

(and some few other discrete devices) to provide an inductive data and power link to the stimulator. To keep the size small it is not only necessary to minimize the number of discrete components but also to pay attention to the size of the signal processing chip. In addition to the implanted stimulator, the system comprises a control unit and a transmitter coil placed on the surface of the skin as shown in fig. 1(a). The implanted stimulator is a small unit placed around a nerve fibre with so-called cuff electrodes to connect the stimulator outputs to the nerve tissue. Fig. 1(b) shows the cuff electrode.

2. SYSTEM SPECIFICATION

Some of the important system requirements are the following: The system must be powered via the inductive link. The stimulation pulses are current pulses which must be programmable in amplitude and duration and each stimulation pulse must be followed by a pulse in the reverse direction ensuring that no charge build-up takes place in the nerve tissue. The maximum amplitude is $2mA$ and the maximum pulse duration is $255\mu s$. The nerve tissue can be expected to exhibit an ohmic resistance of up to $5k\Omega$, implying that the stimulator must be able to generate stimulation pulses of up to $10V$ amplitude. It must be possible to direct the stimulation pulse to a selected electrode among a total number of at least 4 electrodes.

To meet these requirements a low power signal processing chip is required with a unidirectional transmission protocol to program the stimulator pulses and a carrier and modulation scheme which ensures a sufficient power supply.

Fig. 2 shows the stimulator components. In addition to the chip and the coil, a tuning capacitor, a rectifier with a filter capacitor and a capacitor for the charge balancing of the stimulation pulses are required. Also, an external zener diode, D_Z, provides a shunt regulation and overvoltage protection for the chip.

3. TRANSMISSION PROTOCOL

For the power and data transfer to the stimulator, an inductive link with a carrier frequency of $5MHz$ has been chosen. This frequency is chosen as a compromise between inductive coupling (best at high frequencies), and small absorption of tissue and simple electronics (best at low frequencies). The carrier signal is rectified and filtered by the external diode D_1 and the capacitor C_{DD} shown in fig. 2. The supply voltage V_{DD} has been selected to $12V$ in order to be able to generate $2mA$ current pulses into a $5k\Omega$ load. The supply voltage is controlled by the external zener diode D_Z. The data signal is encoded by a pulse amplitude modulation of the carrier with on-off keying. The data rate has been selected to $100kbits/s$. This modulation has been selected because of its simplicity and robustness against imperfections in the transmission but it has the disadvantage that no clock or power is transmitted during the off keying. Hence, the clock recovery circuit must be able to maintain a stable clock frequency during the off keying intervals. For the control of pulse length and amplitude 8-bit words are chosen. Also, to select the output channel for the stimulation, an 8-bit word is chosen. This means that the same transmission protocol can be employed for stimulators with up to 255 output channels. The complete command word to the stimulator is as shown in fig. 3. In addition to the channel select, amplitude and duration control there are two start bits and a cyclic redundancy check word (CRC) for error detection.

4. THE CHIP BLOCK DIAGRAM

A simplified block diagram of the chip is shown in fig. 4. The chip comprises a number of analog circuits functions in addition to a digital control circuit and a digital to analog converter for the pulse amplitude control.

4.1. Voltage regulator

The voltage regulator contains a bandgap reference which is used to control a series regulator for a $3.3V$ supply to the digital blocks in the chip. The main reason for introducing the series regulator is to save power. The power consumption of CMOS logic increases with the square of the supply voltage. At $3.3V$ the CMOS logic is expected to draw about $80\mu A$ leading to a power consumption of about $0.26mW$. With a $12V$ power supply the consumption would be about $3.4mW$, so the series regulator provides a significant power saving even though about $0.7mW$ is dissipated in the series pass transistor of the regulator. The series regulator is fairly conventional.

4.2. Input circuit and clock recovery

The input circuit serves the purpose of extracting the carrier for the generation of the system clock and detecting the modulation for retrieving the data transmitted to the stimulator. Basically, the clock is derived directly from the input signal by taking the signal through a clipping source follower stage, followed by a couple of inverters. The clock derived in this way is of course only detected when the amplitude of the input signal is high, corresponding to a logic 1. When the input is low, no clock signal can be detected since the modulation scheme is a simple on/off keying.

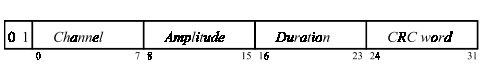


Figure 3. The command word

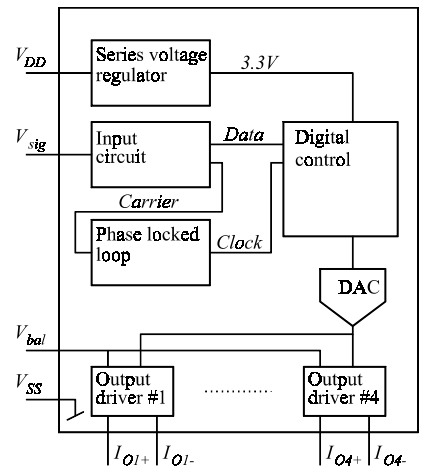


Figure 4. Chip block diagram

4.3. Envelope detection

The envelope detection is done by means of the charge pump circuit shown in fig. 5. The capacitor C_{LP} is continuously discharged by the constant current source I_3. When the carrier is present (i.e. a logic 1 is transmitted), C_{LP} is charged through $M4$ and the current source I_4. The capacitor voltage is at equilibrium when the carrier duty cycle is equal to I_3/I_4 or 1/4. When the duty cycle is higher, as when the input signal is near maximum amplitude, the capacitor voltage tends to $3.3V$ (the digital supply voltage). Conversely, when the duty cycle is lower or even zero, the capacitor voltage tends to V_{SS}. A ripple of about $\Delta V = I_3/(2C_{LP}f_{carrier})$ will be present on the capacitor when the voltage is high due to the discharge when the clock is low. By selecting proper current levels ($I_3 = 2\mu A$) and capacitor size ($1pF$) the ripple can be kept very small and it is filtered out by the current limited inverter $M5 - M6$ and the following standard digital inverter.

4.4. Phase-locked loop

The main purpose of the phase-locked loop is to provide a stable clock for the chip, and to assist in the demodulation of command

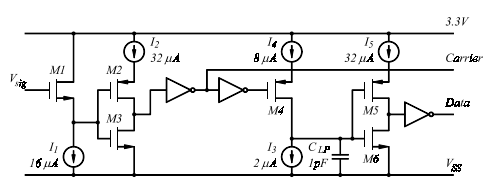


Figure 5. The data input stage

word transmissions. The loop locks onto the carrier frequency during the unmodulated intervals between pulses, and during logic '1' transmissions. During logic '0' transmissions, the frequency of the loop is fixed. This is done by using a carrier detect signal to enable the frequency acquisiontion of the loop. The output frequency of the loop is divided by 5 to provide the $1MHz$ system clock.

The PLL is constructed in a traditional manner, with a phase-frequency detector (PFD), a charge pump filter and a G_m–C voltage controlled oscillator (VCO). A simplified diagram is shown in fig. 6. The VCO is of a type which is suitable for low-power operation at the low MHz frequencies used is this system, and features well-controlled center and operating frequencies [6].

The loop filter is a charge pump, implemented as a passive filter impedance driven by two switched current sources with opposite signs. A detailed analysis of this type of loop filter can be found in [7]. The filter impedance is of second order, making a 3rd-order loop. The main passive component of the filter is a relatively large capacitor, whose value controls the transient behavior of the loop. Simulations indicated that a suitable capacitor size is $100-300pF$, which is quite large but still integrable. An external capacitor was however used, to permit measurements with varying values.

The lock range of the loop is equal to the frequency range of the VCO, which was set to 2.5–10.0MHz. The loop draws an average current of $50\mu A$ from the $3.3V$ digital supply, when locked on a $5MHz$ input signal. The largest part of the supply current is used by the VCO.

The frequency memory feature of the PLL is implemented by adding an enable signal to the current sources in the charge pump. When the pump is disabled, the VCO control voltage is constant, effectively fixing the frequency.

4.5. Amplitude control and output driver

The pulse amplitude is controlled by an 8-bit command as shown in fig. 3. This means that it can be selected to one of 255 current levels ranging from $0-2mA$. The length of the pulse is also selected by an 8-bit command to be $1-255\mu s$. Each pulse is followed by a charge balancing pulse with an amplitude of $128\mu A$ and an appropriate length to achieve the charge balance.

A very simple way to achieve charge balance would be to use an ac-coupling (series capacitor) to the contacting electrodes. However, this would require an external capacitor for each of the output channels which is undesirable because of the space constraints of the stimulator. An alternative is to calculate the charge in the stimulation pulse and derive the pulse length of the charge balancing pulse from this calculation. This can be done in the digital domain from the information received about the pulse amplitude and length, but the calculation would require a good deal of (area consuming) logic and the charge balancing pulse would be quantized in minimum steps of $128\mu A \times 1\mu s$. As an alternative, an analog calculation described below has been selected.

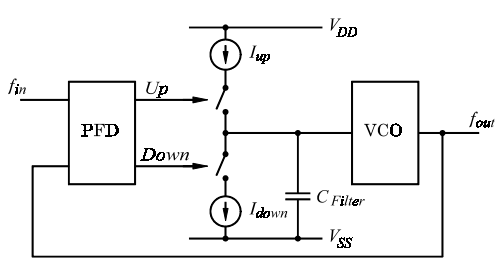


Figure 6. The phase-locked loop

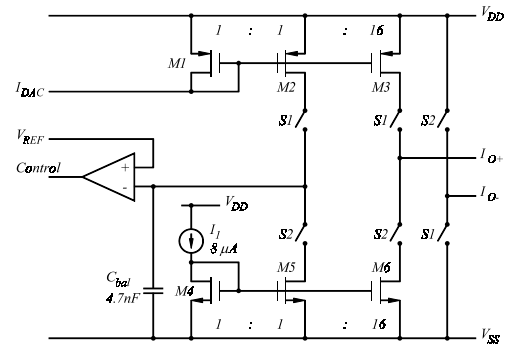


Figure 7. The output driver stage

A simplified schematic of an output driver stage is shown in fig. 7. The switches $S1$ are 'on' when the output driver is selected and a pulse is activated. The switches $S2$ are 'on' when a charge balancing pulse is activated. Basically, the primary pulse is generated by the current mirror $M1, M3$. The input current is derived from a digital to analog converter with an output current range of $0-128\mu A$. The current source I_1 is used for the charge balancing pulse. During the output pulse, $M2$ charges the capacitor C_{bal} with a current equal to the output current divided by 16. During the discharge pulse, the capacitor C_{bal} is discharged with the discharge current divided by 16, i.e. $8\mu A$. Thus, charge balance is obtained when the capacitor voltage is returned (by I_{D5}) to its initial, precharged value detected by a comparator. The comparator output is used by the control logic to terminate the switch control signal S_2. In the actual implementation, the current mirrors $M1-M3$ and $M4-M6$ are implemented as cascode current mirrors to improve the accuracy. The charge balancing also depends on the matching in the output current mirrors. The maximum charge delivered by an output pulse is $Q_{max} = t_{max} \times I_{max} = 255\mu s \times 2mA \simeq 0.5\mu C$. With a mismatch on the order of 5% a charge error of about $25nC$ can be expected. This can be compared to the generally accepted levels of charge accumulation of $1-3\mu C/mm^2$ reported in [4] and [5]. The charge balancing capacitor C_{bal} has a value of $4.7nF$ which is certainly not integratable. However, a single external capacitor can be used for all the output channels.

The digital to analog converter must provide a digitally controllable current output in the range $0-128\mu A$. It is implemented as an array of equally sized current sources controlled by the digital command word.

4.6. Digital control

The digital control logic comprises all the logic needed to decode the command word shown in fig. 3 to the appropriate inputs to the analog to digital converter, the output channel decoder, and the timing of the output pulse. Also, the digital control performs the CRC-8 check of the command word. This reduces the risk

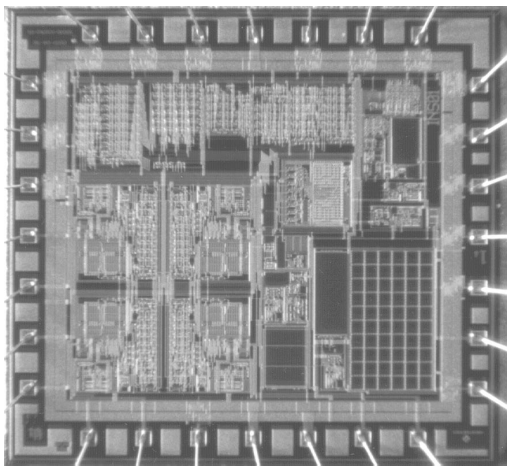


Figure 8. Chip photo of the stimulator chip

of issuing false stimulation pulses. No error correction has been implemented since the application would normally allow a simple retriggering of the stimulation pulses if a pulse was missing due to a transmission error.

5. EXPERIMENTAL RESULTS

Prototype chips have been fabricated in MIETEC's $2\mu m$ CMOS technology. The process has been selected primarily because of its capability to handle the rather large supply voltage. The total chip size is $3mm \times 4mm$. For evaluation purposes the chip has been encapsulated in a standard dual-in-line package. A chip photo is shown in fig. 8. For devices to be used for implantation an alternative encapsulation and mounting together with the discrete devices must be employed.

The chip has been tested with respect to some critical parameters and with respect to functionality. All the circuit blocks have been found to work satisfactorily. The carrier recovery and envelope detection work as predicted. The PLL is able to lock on the carrier and hold its frequency when the carrier is off with a decay of less than $0.1\%/s$. The digital supply voltage has been measured

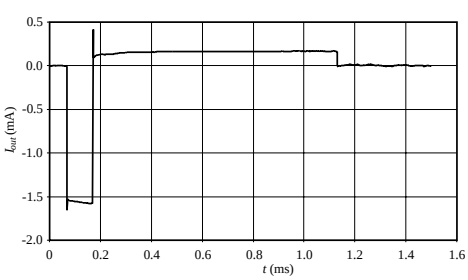


Figure 9. Measured output current pulse from the stimulator

to be within a tolerance of 10%. The total supply current is about $600\mu A$ which is close to the simulated value. The charge balancing circuit shows a mismatch of 5–10% of the stimulation charge, slightly more than expected, but still acceptable.

Functionally, the chip works as expected. Fig. 9 shows the waveform of the output current when the output has been programmed to a duration of $100\mu s$ and an amplitude of $1.5mA$. The output has been measured as the voltage across a $2k\Omega$ load resistor. The charge balancing pulse following the stimulation pulse is clearly seen.

6. CONCLUSION

A prototype chip for a neural stimulator has been designed and fabricated. The chip comprises all the signal processing functions required for the application. A novel method for charge balancing of the stimulation pulses has been developed. Initial measurements show full functionality of the chip and specifications well within the tolerances expected. The next step will be to employ the chip in experiments with functional electrical stimulation of rabbits in order validate that the performance specification of the chip are adequate for the planned applications.

Acknowledgement

The Siemens Foundation is acknowledged for providing financial support for the fabrication of the chip prototypes.

REFERENCES

[1] M. Sawan, S. Robin, B. Provost, Y. Eid, and K. Arabi, "A Wireless Implantable Electrical Stimulator Based on Two FPGAs," *Proc. Third IEEE International Conf. on Electronics, Circuits and Systems*, pp. 1092–95, Rodos, Greece, October 1996.

[2] B. Smith, P. H. Peckham, M. W. Keith, and D. D. Roscoe, "An Externally Powered, Multichannel, Implantable Stimulator for Versatile Control of Paralyzed Muscle," *IEEE Trans. Biomed. Eng.*, Vol. BME-34, pp. 499–598, July 1987.

[3] Gunnar Gudnason, *An Implantable Multichannel Neural Stimulator for Functional Electrical Stimulation*, M. Sc. thesis, Department of Information Technology, Technical University of Denmark, 1997.

[4] N. de N. Donaldson and P. E. K. Donaldson, "When are Actively Balanced Biphasic ('Lilly') Stimulating Pulses Necessary in a Neurological Prosthesis?" *Medical and Biological Engineering and Computing*, Vol. 24, pp. 41–56, January 1996.

[5] T. L. Rose and L. S. Robblee, "Electrical Stimulation with Pt Electrodes. VIII. Electrochemically Safe Charge Injection Limits with 0.2ms Pulses," *IEEE Trans. Biomed. Eng.*, Vol. BME-37, pp. 1118–1120, November 1990.

[6] B. Keeth, R.J. Baker, and H.W. Li, "CMOS Transconductor VCO with Adjustable Operating and Center Frequencies," *Electronics Letters*, Vol. 31, pp. 1397–1398, August 1995.

[7] Floyd M. Gardner, "Charge-Pump Phase-Lock Loops," *IEEE Transactions on Communications*, Vol. COM-28, pp. 1849–1858, November 1980.

Analog Integrated Circuits and Signal Processing, 22, 81–89 (1999)

A Chip for an Implantable Neural Stimulator

GUNNAR GUDNASON AND ERIK BRUUN

Department of Information Technology, Building 344, DTU, DK-2800 Lyngby
E-mail: gg@it.dtu.dk, bruun@it.dtu.dk

MORTEN HAUGLAND

Center for Sensory-Motor Interaction, Aalborg University, Fr. Bajersvej 7D, DK-9220 Aalborg
E-mail: mh@miba.auc.dk

Received January 18, 1999; Revised May 12, 1999; Accepted May 13, 1999

Abstract. This paper describes a chip for a multichannel neural stimulator for functional electrical stimulation (FES). The purpose of FES is to restore muscular control in disabled patients. The chip performs all the signal processing required in an implanted neural stimulator. The power and digital data transmission to the stimulator passes through a 5 MHz inductive link. From the signals transmitted to the stimulator, the chip is able to generate charge-balanced current pulses with a controllable length up to 256 μs and an amplitude up to 2 mA, for stimulation of nerve fibers. The quiescent current consumption of the chip is approx. 650 μA at supply voltages of 6–12 V, and its size is $3.9 \times 3.5\,\text{mm}^2$. It has 4 output channels for use in a multipolar cuff electrode.

Key Words: neural stimulator, inductive link, implantable microsystems, FES

1. Introduction

Functional electrical stimulation is the activation of physical functions through electrical stimulation of nerve tissue. It can be applied for instance to victims of spinal chord injuries who have lost control over part of their body (paralysis) or to patients suffering from foot drop. By stimulating the nerves controlling the muscles or the muscle tissue directly, some degree of muscular control is regained. Several implantable solutions to this problem have been described in the literature, each with their distinctive features.

The system presented here is designed to be attached directly to a nerve, and to provide fine control of the muscle at the end of the nerve through the use of several stimulation channels. Other systems have been reported in the literature for injection by hypodermic needle into muscle tissue [1–4], and they have typically one or a few stimulation channels. A very important requirement for injectable stimulators is clearly a very small size. Multiple muscle sites are stimulated by using a number of such devices. On the other hand, stimulators have been presented with one central processing unit, and wire connections out to the stimulation sites [5]. One advantage of such an arrangement is that the stimulator can be placed just below the skin, permitting efficient power and data transfer.

A common feature of most implantable stimulators is that each single stimulation pulse corresponds to a command transmitted over the inductive link. The very large number of channels (up to several hundred) used in intracortical stimulating electrode arrays [6] requires a different type of solution. The high data rate necessary to specify all stimulation pulses is reduced through use of on-chip pulse timers and pulse parameter control [7].

The stimulator is placed at the stimulation site. Thus, no implanted wiring is required, yet the stimulator is capable of providing selective multi-channel stimulation. In the prototype described in this paper, 4 channels are available but the design can easily be modified to include more channels without excessive increase in power consumption but at the expense of more chip area.

The stimulator is designed for use with a

multipolar electrode connected to a nerve trunk, and to provoke different modes of motion in the attached musculature by exciting different nerve fibers. Fine control over the pulse parameters is necessary to achieve this goal. The chip produces asymmetric biphasic current pulses, which calls for an on-chip calculation of the second pulse phase's duration. This problem was solved by a novel analog calculation method.

2. System Overview and Specification

For the stimulator described here, an important consideration is the physical size and the option of having several output channels. The stimulator is an implantable unit comprising electrodes to contact the nerves, a signal processing chip to generate the stimulation pulses, and a coil (and some few other discrete devices) to provide an inductive data and power link to the stimulator. To keep the size small, it is not only necessary to minimize the number of discrete components but also to pay attention to the size of the signal processing chip. In addition to the implanted stimulator, the system consists of a control unit and a transmitter coil placed on the surface of the skin as shown in Fig. 1(a). The implanted stimulator is a small unit placed around a nerve fiber with so-called cuff electrodes to connect the stimulator outputs to the nerve tissue. Cuff electrodes are suitable for selective stimulation of relatively large nerves [9,10], such as the peroneal nerve in the lower leg. Fig. 1(b) shows the cuff electrode.

Some of the important system requirements are the following: The system must be powered via the

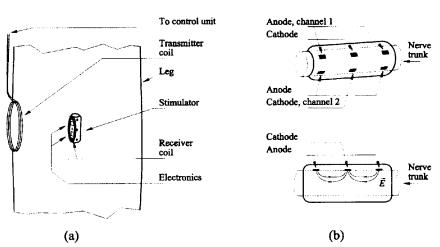


Fig. 1. (a) Placement of implanted stimulator and transmitter coil on the lower leg. (b) Cuff electrode (perspective view and electric field during stimulation).

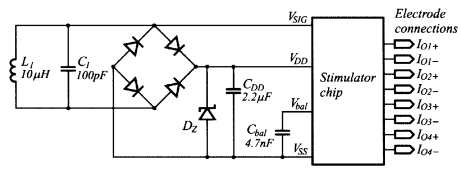


Fig. 2. The stimulator components.

inductive link. The stimulation pulses are rectangular current pulses which must be programmable in amplitude and duration and each stimulation pulse must be followed by a pulse in the reverse direction, ensuring that no charge build-up takes place in the nerve tissue. The maximum pulse amplitude is $I_1 = 2\,\text{mA}$ and the maximum pulse duration is $t_1 = 256\,\mu\text{s}$. The discharge pulse has a fixed amplitude $I_2 = 128\,\mu\text{A}$ and a duration t_2 which is adjusted by the chip for zero total charge. The nerve tissue can be expected to exhibit an ohmic resistance of up to $5\,\text{k}\Omega$, implying that the stimulator must be able to generate stimulation pulses of up to 10 V amplitude. It must be possible to direct the stimulation pulse to a selected electrode among a total number of at least 4 electrodes. Unused electrode terminals must be floating between pulses, to avoid interfering with the current pattern set up by an active electrode.

To meet these requirements a low power signal processing chip is required with a unidirectional transmission protocol to program the stimulator pulses and a carrier and modulation scheme which ensures a sufficient power supply.

Fig. 2 shows the stimulator components. In addition to the chip and the coil, a tuning capacitor, a rectifier with a filter capacitor and a capacitor for the charge balancing of the stimulation pulses are required. Also, an external zener diode, D_Z, provides a shunt regulation and overvoltage protection for the chip.

3. Transmission Protocol

For the power and data transfer to the stimulator, an inductive link with a carrier frequency of 5 MHz has been chosen. This frequency is chosen as a compromise between inductive coupling (best at high frequencies), and small absorption of tissue and simple electronics (best at low frequencies). The

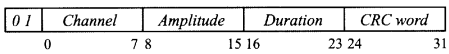


Fig. 3. The command word.

carrier signal is rectified and filtered by the external bridge rectifier and the capacitor C_{DD} shown in Fig. 2. The supply voltage V_{DD} has been selected to 12 V in order to be able to generate 2 mA current pulses into a 5 kΩ load. The data signal is encoded by a pulse amplitude modulation of the carrier with on-off keying. The data rate has been set to 100 kbits/s. This modulation type has been selected because of its simplicity and robustness against imperfections in the transmission, but it has the disadvantage that no clock or power is transmitted during the off keying. Hence, the clock recovery circuit must be able to maintain a stable clock frequency during the off keying intervals. 8-bit words are chosen to represent the pulse amplitude and duration, and to select the output channel for stimulation. This means that the same transmission protocol can be employed for stimulators with up to 256 output channels. The complete command word to the stimulator is as shown in Fig. 3. In addition to the channel select, amplitude and duration control there are two start bits and a cyclic redundancy check word (CRC) for error detection.

4. The Chip Block Diagram

A simplified block diagram of the chip is shown in Fig. 4. The chip comprises a number of analog circuits functions in addition to a digital control circuit and a digital to analog converter for the pulse amplitude control.

4.1. *Voltage Regulator*

The voltage regulator contains a bandgap reference which is used to control a series regulator for a 3.3 V supply to the digital blocks in the chip. Fig. 5 shows the voltage regulator. The operation of this circuit is relatively simple, and can be described as follows. The two feedback resistors provide both a minimum bias current for the pass transistor, and an input to the opamp which is equal to 4/11 times the output voltage. A pass transistor is added at the output of the

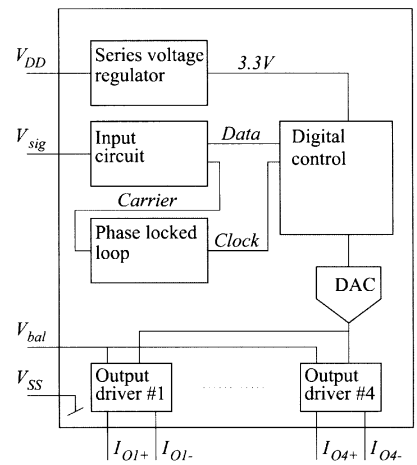


Fig. 4. Chip block diagram.

operational amplifier to provide adequate current sourcing capability. The addition of a capacitor at the regulator output is necessary because of the current transients drawn by the digital load. With a 300 pF capacitor, the digital supply voltage transients are smaller than 10%.

The main reason for introducing the series regulator is to save power. The power consumption of CMOS logic increases with the square of the supply voltage. At 3.3 V, the CMOS logic is expected to draw about 80 μA, leading to a power consumption of about 0.26 mW. With a 12 V power supply the consumption would be about 3.4 mW, so the series regulator provides a significant power saving even though about 0.7 mW is dissipated in the series pass transistor of the regulator.

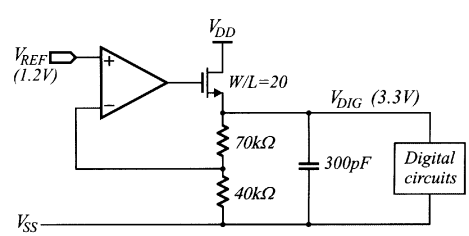


Fig. 5. The 3.3 V digital supply regulator.

4.2. *Input Circuit and Clock Recovery*

The input circuit serves the purpose of extracting the carrier for the generation of the system clock and detecting the modulation for retrieving the data transmitted to the stimulator. Basically, the clock is derived directly from the input signal by taking the signal through a clipping source follower stage, followed by a pair of inverters. The clock derived in this way is of course only detected when the amplitude of the input signal is high, corresponding to a logic 1. When the input is low, no clock signal can be detected since the modulation scheme is a simple on/off keying. To provide a continuous clock signal to the chip, a phase-locked loop is employed. The loop is a conventional digital PLL, with a phase-frequency detector, a charge pump driving a passive filter impedance and a VCO [11]. The loop locks onto the carrier frequency during logic 1 transmissions. During logic 0 transmissions the VCO of the PLL holds the frequency. This is achieved by using a carrier detect signal to enable the charge pump in the PLL. Only when a carrier is detected will the PLL try to follow the input frequency. The output from the PLL is divided by 5 to provide a system clock of 1 MHz.

The input signal conditioning and envelope detection is done by means of the charge pump circuit shown in Fig. 6. The gate of the source follower $M1$ is connected directly to the off-chip signal V_{SIG}, which can vary from one diode drop below V_{SS} to one diode drop above V_{DD} (up to 12 V). The range of possible input voltages is determined by the external bridge rectifier (Fig. 2). The current-limited inverter $M2$-$M3$ and the following standard inverter clean up the clipped input signal, so the recovered *carrier* signal has sharp transitions.

The envelope detection is performed by low-pass filtering of the *carrier* signal. The capacitor C_{LP} is continuously discharged by the constant current source I_3. When the *carrier* signal is high, C_{LP} is

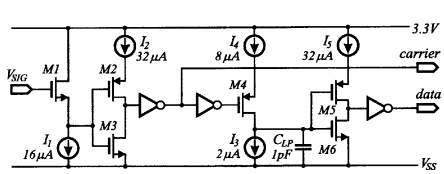


Fig. 6. The signal input and conditioning stage.

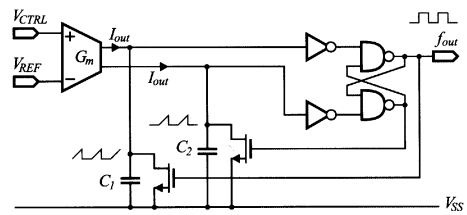


Fig. 7. The VCO used in the PLL.

charged through $M4$ and the current source I_4. The capacitor voltage is at equilibrium when the *carrier* duty cycle is equal to I_3/I_4 or 1/4. When the duty cycle is higher, as when the input signal is near maximum amplitude, the capacitor voltage tends to 3.3 V (the digital supply voltage). Conversely, when the duty cycle is lower or even zero, the capacitor voltage tends to V_{SS}. A ripple of about $\Delta V = I_3/(2C_{LP}f_{carrier})$ will be present on the capacitor when the voltage is high, due to the discharge when *carrier* is low. By selecting a proper capacitor size (1 pF) and current levels ($I_3 = 2\,\mu$A), the ripple can be kept very small. The ripple is filtered out by the current limited inverter $M5$–$M6$ and the following standard inverter.

The VCO is shown in Fig. 7. This is a G_m-C circuit reported in [12], which is suitable for low-power operation at moderate frequencies. The two transconductor outputs are used to alternately charge two identical capacitors to the threshold voltages of the following inverters. The flip-flop controls the switch transistors connected across the capacitors, and thus determines the charging sequence. The frequency range of the VCO is determined by the capacitor values and the output current range of the transconductor. A frequency range of 2.5–10 MHz was chosen to ensure that the PLL could lock onto the 5 MHz input signal, despite worst-case process variations.

4.3. *Amplitude Control and Output Driver*

The pulse amplitude is controlled by an 8-bit command as shown in Fig. 3. This means that it can be selected to one of 255 current levels ranging from 0–2 mA. The duration of the pulse is also selected by an 8-bit command to be 1–256 μs. Each pulse is followed by a charge balancing pulse with an

amplitude of 128 μA and an appropriate duration to achieve the charge balance.

A very simple way to achieve charge balance would be to use an AC coupling (series capacitor) to the contacting electrodes. However, this would require an external capacitor for each of the output channels, which is undesirable because of the space constraints of the stimulator.

An alternative is to calculate the charge in the stimulation pulse and derive the duration of the discharge pulse from this calculation. This can be done in the digital domain from the information received about the pulse amplitude and duration, as $t_2 = t_1 I_1 / I_2$. The amplitude I_1 is quantized in steps of $I_q = 8\,\mu$A, so $I_1 = ampl \times I_q$ where *ampl* is the 8-bit amplitude data. The amplitude of the discharge pulse is fixed at $I_2 = 128\,\mu\text{A} = 16 I_q$. The duration of the discharge pulse can then be expressed as

$$t_2 = (1\mu\text{s}) \times \left\lceil \frac{ampl \times dur}{16} \right\rceil \qquad (1)$$

where *dur* is the 8-bit pulse duration data. This calculation requires one multiplication of 8-bit operands for a 16-bit result and a subsequent shift by 4 bits. The shift can be omitted by simply using the highest 12 bits of the result for the pulse length.

While this approach is conceptually simple, it requires the addition of a multiplier to the circuit, and extensive additions to the control circuitry. The 2 μm technology used for this chip does not lend itself well to compact digital circuits, so the area penalty is appreciable. On the other hand, the calculation of the pulse length requires only a small amount of energy per pulse. The discharge pulse is quantized in steps of 128 μA × 1 μs. The quantization error is smaller than the mismatch between the current sources, for all but the smallest stimulation pulses.

After consideration of complexity, area and power, the analog calculation method described below was selected. It is based on integration of the current in both pulse phases, and termination of the pulse when the integral reaches zero. This method has the disadvantage of requiring an external capacitor.

A simplified schematic of an output driver stage is shown in Fig. 8. The switches *S*1 are "on" when the output driver is selected and a stimulation pulse is activated. The switches *S*2 are "on" when a discharge pulse is activated. Basically, the primary pulse is generated by the current mirror *M*1, *M*3. The input

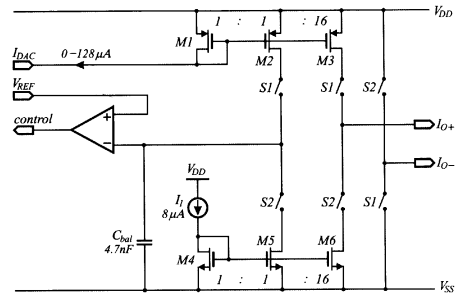


Fig. 8. The output driver stage.

current is derived from a digital to analog converter with an output current range of 0–128 μA. The current source I_1 is used for the discharge pulse. During the output pulse, *M*2 charges the capacitor C_{bal} with a current equal to the output current divided by 16. During the discharge pulse, the capacitor C_{bal} is discharged with the discharge current divided by 16, i.e. 8 μA. Thus, charge balance is obtained when the capacitor voltage is returned (by I_{D5}) to its initial, precharged value detected by a comparator. The comparator output is used by the control logic to terminate the switch control signal S_2. In the actual implementation, the current mirrors *M*1–*M*3 and *M*4–*M*6 are implemented as cascode current mirrors to improve the accuracy. The charge balancing also depends on the matching in the output current mirrors. The maximum charge delivered by an output pulse is $Q_{max} = t_{max} \times I_{max} = 256\,\mu\text{s} \times 2\,\text{mA} \simeq 0.5\,\mu\text{C}$. With a mismatch on the order of 5%, a charge error of about 25 nC can be expected. This can be compared to the generally accepted levels of charge accumulation of 1–3 μC/mm^2 reported in [14] and [15]. The charge balancing capacitor C_{bal} has a value of 4.7 nF, which is certainly not integratable. However, a single external capacitor can be used for all the output channels.

4.4. Digital to Analog Converter

The digital to analog converter translates the 8-bit amplitude command into a current in the range 0–128 μA. This current is used by the output driver, which amplifies it by a factor of 16 before sending it through the electrode connection, as shown in Fig. 8. The DAC is a simple 8-bit binary-weighted current

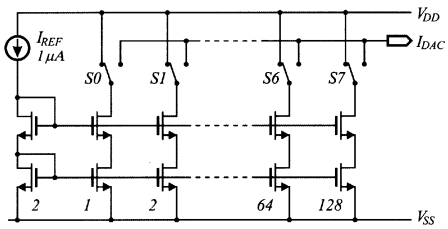


Fig. 9. The digital to analog converter. The numbers show the number of unit transistors.

source type, as shown in Fig. 9. Unit transistors were used for the current sources, and the current source corresponding to bit i contains 2^i such transistors. The transistors corresponding to each bit were laid out with a common centroid to improve matching. Cascode transistors were added to each current source.

Since the DAC has several microseconds available to perform each conversion, high speed is not an issue. In order to save power, the DAC is only turned on during a stimulation pulse.

4.5. *Digital Control*

The digital control logic comprises all the logic needed to decode the command word shown in Fig. 3 to the appropriate inputs to the analog to digital converter, the output channel decoder, and the timing of the output pulse. Also, the digital control performs the CRC-8 check of the command word. This check reduces the risk of issuing false stimulation pulses. No error correction has been implemented since the application would normally allow a simple retriggering of the stimulation pulses if a pulse was missing due to a transmission error.

4.6. *Area and Current Consumption*

Since size and power consumption are among the main figures of merit for the stimulator, it is interesting to note these attributes for individual blocks. Table 1 shows the simulated current consumption (at a supply voltage of $V_{DD} = 10\,\text{V}$) and the layout area for the blocks in the block diagram (Fig. 4). Since it is not possible to measure the current consumption of each block separately, the figures in the table are simulation values. They are therefore not very accurate, especially for the digital circuits.

Table 1. The area and quiescent current consumption of the main stimulator components.

Circuit	$I_{DD}(\mu\text{A})$	Area(mm^2)
3.3 V regulator	55	1.3
Input circuit	30	0.06
PLL	50	0.42
Output drv. and DAC	110	3.1
Digital logic	80	1.8
Voltage ref. and bias	60	0.12
Support circuits	130	0.28
Total	515	

These are quiescent supply current figures. During a stimulation pulse the supply current rises by approximately $130\,\mu\text{A} + 1.1$ times the pulse current. The support circuits mentioned in the table include a supply voltage sensitive global reset stage and test structures.

5. Experimental Results

Prototype chips have been fabricated in the Alcatel Microelectronics 2 μm CMOS technology, available through the EUROPRACTICE IC service. The process has been selected primarily because of its ability to handle a supply voltage of 12 V. The total chip size is $3.9 \times 3.5\,\text{mm}^2$. For evaluation purposes the chip has been encapsulated in standard DIL and SOIC packages. A chip photo is shown in Fig. 10.

The chip has been tested both with and without the inductive link in the signal path. Most quantitative results below were obtained with a signal generator and power supply connected directly to the chip terminals, but the functionality of the chip has been verified with an inductive link.

All the circuit blocks have been found to work satisfactorily. The carrier recovery and envelope detection work as predicted. The PLL is able to lock onto the carrier and hold its frequency when the carrier is off with a decay of less than 0.1%/s. The digital supply voltage has been measured to be within a tolerance of 10%.

Due to the limited capacity of the power link to the implant, the supply current of the chip was kept at a minimum. The current consumption of the chip is

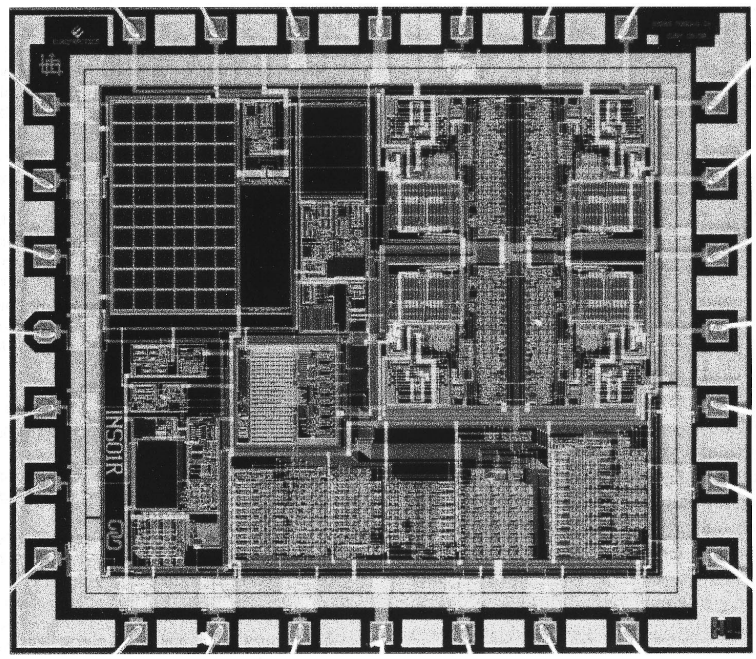


Fig. 10. A photograph of the stimulator chip.

shown in Fig. 11, in the absence of stimulation commands. During output pulses, the supply current rises by an amount corresponding to the pulse amplitude. In the useful supply voltage interval $V_{DD} = 6-12\,\mathrm{V}$, the supply current is 600–750 μA,

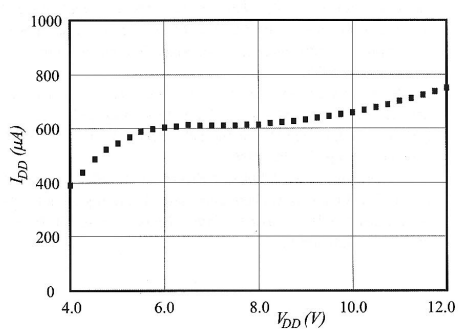


Fig. 11. The measured quiescent supply current of the chip.

which is close to the simulated value. 12 V is the maximum supply voltage specified for this technology, and operating the chip near this limit may degrade the life expectancy of the stimulator. Should the electrode resistance prove to be smaller than the specified maximum of 5 kΩ, the supply voltage can be set to a lower level.

Functionally, the chip works as expected. Fig. 12 shows the waveform of the output current when the output has been programmed to a duration of 100 μs and an amplitude of 1.5 mA. The output has been measured as the voltage across a 2 kΩ load resistor. The charge balancing pulse following the stimulation pulse is clearly seen.

Tests were performed with the inductive link at 5 MHz, to verify the functionality of the chip under realistic electrical conditions. The dimensions of the coils depend on the actual placement of the implant, but as an approximation the following coils were used for testing. The transmitter coil has a diameter of 50 mm, 10 turns of 0.5 mm copper wire, and an

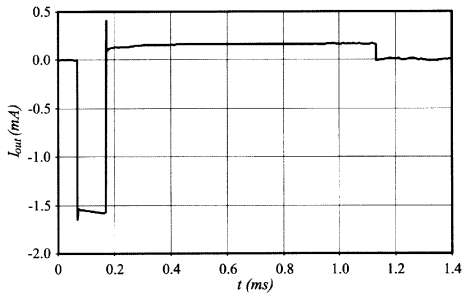


Fig. 12. A measured output current pulse from the stimulator.

inductance of 10 μH. The receiver coil has an area of 2.5 × 14 mm^2, 25 turns of 0.075 mm copper wire, and an inductance of 10 μH. With the transmitter currently in use, the maximum range is approximately 40 mm.

The standard packages used for the prototype chip are obviously not suitable for implantation, both because of their bulk and their poor protection against the hostile environment. A packaging solution is being examined where the chip is mounted together with the discrete components onto a common substrate, and subsequently encapsulated in a polymer and silicone.

6. Conclusion

A prototype chip for a neural stimulator has been designed and fabricated. The chip comprises all the signal processing functions required for the application. A novel method for charge balancing of the stimulation pulses has been developed. Initial measurements show full functionality of the chip and specifications well within the tolerances expected. The next step will be to employ the chip in experiments with functional electrical stimulation of rabbits in order to validate that the performance specifications of the chip are adequate for the planned applications.

Acknowledgments

The Siemens Foundation is acknowledged for providing financial support for the fabrication of the chip prototypes.

References

1. G. E. Loeb, C. J. Zamin, J. H. Schulman, and P. A. Troyk, "Injectable stimulator for functional electrical stimulation." *Med. & Biol. Eng. & Comput.* 29, pp. NS13–NS19, 1991.
2. M. Nardin and K. Najafi, "A multichannel neuromuscular microstimulator with bi-directional telemetry." *1995 Int. Conf. on Solid-State Sensors and Actuators, Transducers '95* Stockholm, pp. 59–62, 1995.
3. T. M. Cameron, G. E. Loeb, R. A. Peck, J. H. Schulman, P. Strojnik, and P. A. Troyk, "Micromodular implants to provide electrical stimulation of paralyzed muscles and limbs." *IEEE Trans. Biomed. Eng.* BME-44, pp. 781–790, 1997.
4. B. Ziaie, M. Nardin, A. R. Coghlan, and K. Najafi, "A single-channel implantable microstimulator for functional neuromuscular stimulation." *IEEE Trans. Biomed. Eng.* BME-44, pp. 909–920, 1997.
5. B. Smith et al., "An externally powered, multichannel, implantable stimulator-telemeter for control of paralyzed muscle." *IEEE Trans. Biomed. Eng.* BME-45, pp. 463–475, 1998.
6. C. Kim and K. D. Wise, "A 64-site multishank CMOS low-profile neural stimulating probe." *IEEE J. Solid-State Circuits* SC-31, pp. 1230–1238, 1996.
7. K. E. Jones and R. A. Normann, "An advanced demultiplexing system for physiological stimulation." *IEEE Trans. Biomed. Eng.* BME-44, pp. 1210–1220, 1997.
8. Z. Tang and P. H. Peckham, "Multichannel implantable stimulation and telemetry system for neuromuscular control." *Proc. 16th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Soc.* 1, pp. 442–443, 1994.
9. G. E. Loeb and R. A. Peck, "Cuff electrodes for chronic stimulation and recording of peripheral nerve activity." *J. of Neurosci. Methods* 64, pp. 95–103, 1996.
10. C. Veraart, W. Grill, and J. T. Mortimer, "Selective control of muscle activation with a multipolar nerve cuff electrode." *IEEE Trans. Biomed. Eng.* BME-40, pp. 640–653, 1993.
11. Floyd M. Gardner, "Charge-pump phase-locked loops." *IEEE Trans. Comm.* COM-28, pp. 1849–1858, 1980.
12. B. Keeth, R. J. Baker, and H. W. Li, "CMOS transconductor VCO with adjustable operating and center frequencies." *Electronics Letters* 31, pp. 1397, 1995.
13. Gunnar Gudnason, *An Implantable Multichannel Neural Stimulator for Functional Electrical Stimulation.* M.Sc. thesis, Department of Information Technology, Technical University of Denmark, 1997.
14. N. de N. Donaldson and P. E. K. Donaldson, "When are actively balanced biphasic ('Lilly') stimulating pulses necessary in a neurological prosthesis?" *Med. & Biol. Eng. & Comput.* 24, pp. 41–56, 1986.
15. T. L. Rose and L. S. Robblee, "Electrical stimulation with Pt electrodes. VIII. Electrochemically safe charge injection limits with 0.2 ms pulses." *IEEE Trans. Biomed. Eng.* BME-37, pp. 1118–1120, 1990.

Morten Haugland was born on June 1, 1964 in Holstebro, Denmark. He graduated as M.Sc.E.E. in 1989 and got a Ph.D. in Biomedical Engineering in 1994 from Aalborg University, Denmark. From 1989 to 1990 he worked as Research Assistant in Dr. J. A. Hoffer's laboratory at the University of Calgary, Canada. He is presently employed as associate professor at the Center for Sensory-Motor Interaction at Aalborg University. His main interests are within design and application of implantable electrodes and eletronics for restoration of motor and sensory function of paralyzed people.

Gunnar Gudnason received the M.Sc. degree in Electrical Engineering in 1997 from the Technical University of Denmark (DTU). He is currently working towards a doctoral degree at the Department of Information Technology, DTU. His main research interest is the design of analog integrated circuits for implantable electronics.

Erik Bruun received the M.Sc. and Ph.D. degrees in electrical engineering in 1974 and 1980, respectively, from the Technical University of Denmark. In 1980 he also received the B.Com. degree from the Copenhagen School of Commerce.

From January 1974 to September 1974 he was with Christian Rovsing A/S, working on the development of space electronics and test equipment for space electronics. From 1974 to 1980 he was with the Laboratory for Semiconductor Technology at the Technical University of Denmark, working in the field of MNOS memory devices, I^2L devices, bipolar analog circuits, and custom integrated circuits. From 1980 to 1984 he was with Christian Rovsing A/S, where he headed the development of custom and semicustom integrated circuits. From 1984 to 1989 he was the managing director of Danmos Microsystems ApS, a company specializing in the development of application specific integrated circuits and in design tools for the electronics industry. Since 1989 he has been a Professor of analog electronics at the Technical University of Denmark. Since 1995 he has served as head of the Sector of Information Technology, Electronics, and Mathematics at the Technical University of Denmark, and he is a member of the senate of the University.

He has published numerous papers on integrated circuits and analog signal processing in international journals and at international conferences and holds many professional positions. He has been guest editor of *IEEE Journal of Solid-State Circuits* and of *Analog Integrated Circuits and Signal Processing* and he has been a member of many conference program committees.

His current research interests are in the areas of current mode analog signal processing, intelligent transducers and integrated circuits for mobile phones.

A low-power ASK demodulator for inductively coupled implantable electronics

Gunnar Gudnason
Dept. of Information Technology, Technical University of Denmark, DK-2800 Lyngby
gg@it.dtu.dk

Abstract

An amplitude shift keying (ASK) demodulator is presented which is suitable for implantable electronic devices that are powered through an inductive link. The demodulator has been tested with carrier frequencies in the range 1–15 MHz, covering most commonly used frequencies. Data rates up to several 100kbit/s are supported, suitable for complex implants such as stimulating electrode arrays or visual implants. The circuit is compatible with modulation depths in the range 10–100%. The low end of the range permits data transmission without significant reduction in power transfer, or the use of transmitter designs with limited modulation capability. The power consumption is 60 µW from a 3 V supply. The circuit has been implemented in a standard CMOS process.

1. Introduction

Inductively coupled links are widely used nowadays in conjunction with high-performance implantable devices, to provide wireless power and data transmission to the implant. The link consists of two resonant RLC circuits, with the external primary circuit driven by a power amplifier, and the implanted secondary circuit acting as an antenna. Several different circuit topologies are possible [1] but it will be assumed here without loss of generality that the secondary circuit is parallel resonant. The power amplifier is usually a switching type such as a class D or E with an amplitude modulated output [2]. Efficient transmitter design and power transfer requirements can limit the AM modulation depth to 10–20%.

One of the special requirements of implantable systems is that maximum power transfer may be desired during periods when the data link is idle, as for example between stimulation commands in a stimulator system. Modulation of the carrier during idle periods is unwanted in such cases, and the idle periods can therefore not be used to establish the modulation levels in the detector.

The operating conditions for implantable electronics are in many respects similar to those experienced by contactless smart card ICs. The main difference is the available power in the transmitter. An ASK method based on sensing the current passed through an on-chip shunt regulator is presented in [3], but in a portable battery-operating system, the ideal situation is when little or no current is wasted in a shunt regulator. Ideally, the supply voltage of the implant is regulated by varying the transmitted power instead. This can be done by monitoring the supply voltage through a bidirectional data link, and regulating the transmitted power accordingly.

The weak coupling commonly found between the primary and secondary coils means that the power transfer is very inefficient. Any reduction in the power consumption of an implant can therefore significantly affect the power economy of the system as a whole, and in the case of battery-operated external transmitters, increase battery life.

One of the basic parts found in all digitally controlled implants is the demodulator, and this paper presents a versatile ASK demodulator, which features a low power consumption without limiting the bit rate or setting unreasonable constraints on the modulation form.

2. System overview

Figure 1 shows a block diagram of the demodulator. The input signal and power are received through an antenna consisting of a tuned parallel LC resonator. The power for the implant is extracted by a bridge rectifier and accumulated in a storage capacitor (not shown). Depending on the technology, the rectifier may be on-chip or external. The RF signal is taken from one side of the antenna and fed through a passive input network to a current-mode squarer. The squarer has a differential output which is connected to a differential low-pass filter, and an extracted carrier output which can be used as a system clock. The output of the LPF is connected to a level detector which extracts the digital modulation data from the envelope information.

2.1. Input network

The amplitude of the antenna signal can be far greater than that which can be handled by modern CMOS technologies. The signal can therefore not be connected directly to a transistor gate or diffusion, but must be conditioned in some way. Input protection diodes are not used since they would interfere with the operation of the bridge rectifier. In some cases it is desirable to maintain a supply voltage external to the chip which is higher than the on-chip supply. In that case, a voltage regulator is inserted into the supply path to the chip in figure 1, so the bridge rectifier does not provide sufficient input protection.

The passive input network consists of a capacitive divider and a resistor R_1 in series with the input resistance R_i of the squarer. This network has two purposes. One is

Reprinted with permission from the proceedings of the ESSCIRC 2000 conference.

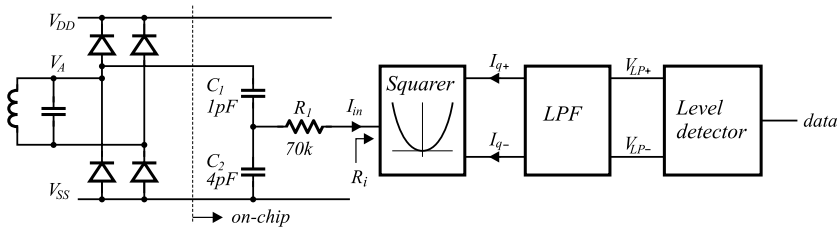


Figure 1. A block diagram of the demodulator circuit.

to convert the input voltage V_A to an input current of suitable amplitude for the squarer, and the other is to reduce the signal voltage seen by the active devices in the circuit. The capacitors are implemented using metal layers, so they can withstand quite high voltages. The capacitors are stacked on top of each other so that the chip substrate is shielded from input signal by the bottom plate of C_2.

Along with the source resistance R_S of the inductive link, the network has a second-order bandpass transfer function. Assuming that the corner frequencies are well separated (a decade or more), the transfer function in the middle of the band is approximately $I_{in}/V_A = C_1/((C_1 + C_2)(R_1 + R_i))$. The lower and upper corner frequencies are $1/\omega_l = (R_1 + R_i)(C_1 + C_2)$ and $1/\omega_u = R_S C_1$ respectively. These must be fitted to the used carrier frequency range.

2.2. Squarer

The current squarer is shown in figure 2. It is based on a four-quadrant class AB multiplier presented in [4], and uses the square-law characteristic of the MOS transistor in strong inversion. The output is given by $(I_{o+} - I_{o-}) = I_{in}^2/8I_b$ where I_b is the bias current of the input transistors.

One of the features which makes this circuit attractive for our purposes is that a relatively high small-signal bandwidth can be achieved with low power levels. Simulations show a bandwidth greater than 30 MHz with a total bias current of 5 µA. The supply current of the squarer is signal-dependent due to the class AB operation, and rises from 5 µA to 10 µA at the maximum specified signal amplitude.

One limitation of the circuit is that it relies on matching of P- and N-type devices to implement the quadratic transfer function. While this is impossible to achieve in practice, it does not matter in this application since the squarer is only used as an envelope detector. The mismatch of the current mirrors, which are implemented with high-swing cascodes, also affects the accuracy of the circuit, but to a much lesser extent.

The two current mirrors have secondary outputs which are connected together. The high-impedance node at these secondary outputs slews to the positive supply or to ground according to the sign of the input current, thus creating a clock signal corresponding to the carrier frequency. Many implantable electronic devices use the signal carrier as a timing reference and to clock sequential logic.

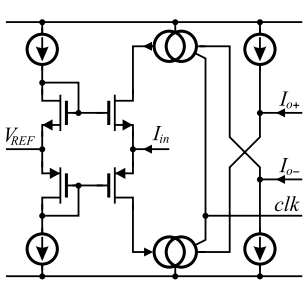


Figure 2. The current-mode squarer circuit. The bias current in each branch is 1 µA.

2.3. Low-pass filter

The low-pass filter is used to extract the envelope data from the squared input signal. It is a differential third-order filter, consisting of a passive first-order section and an active section which is implemented as a G_m–C filter. Figure 3 shows the filter.

The cutoff frequency of the filter is set with respect to the highest intended modulation frequency, taking process variations into account. In this case it was set to a nominal frequency of 350 kHz, to permit unencoded data rates of up to approximately 200kbit/s.

Figure 4 shows the transconductor used in the filter implementation, along with the common-mode feedback circuit which is necessary for each of the two internal differential nodes of the filter. The CMFB circuit is a current-

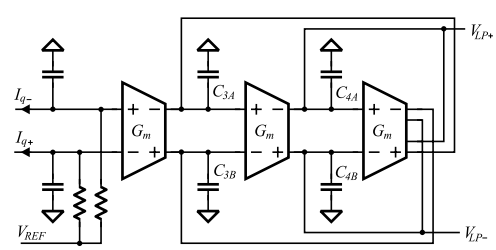


Figure 3. The differential LPF implementation.

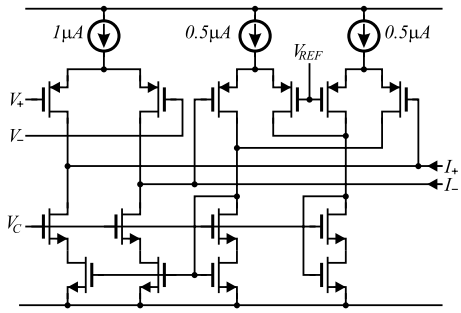


Figure 4. A single-output filter transconductor, with CMFB circuit.

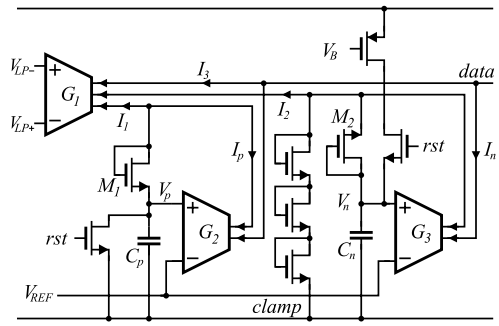


Figure 5. The peak detector and data signal extraction circuit.

steering type which causes relatively little distortion of the differential mode signal. The transconductor itself is based on a simple differential pair. The nonlinearity of the transconductor is of little importance at the applied signal level (up to 200 mV differential). The 3dB-frequency of the transconductors is approximately 10 MHz, which is sufficiently high not to affect the transfer function of the filter.

2.4. Level detector

The power requirements of the implant and considerations of the efficiency of the link generally require that the transmitter be operated at a constant level while the data link is idle. For much the same reason it is desirable to eliminate long synchronization/start sequences before each transmission. The ASK modulation levels are therefore not known by the demodulator *a priori*, and must be established quickly to detect the transmitted data. This eliminates the possibility of a LP filter with a long time constant to determine the average signal level. A method commonly used in sophisticated receivers, like the digital conversion and storage of the transmitted envelope with *a posteriori* processing, is not compatible with the power consumption criteria.

The level detector presented here uses only a pair of start bits for a simple baseband modulation scheme, or a single start bit in the case of biphase ("Manchester") encoding, to identify the two modulation levels and the threshold level.

A diagram of the detector is shown in figure 5. It uses three single-ended transconductors with multiple outputs. All three transconductors have the same transconductance G, except for the third output of G_1 which has twice the magnitude of the other outputs, $I_3 = 2I_2 = 2I_1$.

The positive peak voltage at the input is stored on C_p (referred to V_{REF}), while the negative peak is stored on C_n. In the following calculation of the transfer function from the input to the positive peak voltage V_p, the diode transistor M_1 can be ignored. The justification for this will be apparent from the results.

From the diagram, we have the small-signal currents $i_1 = G(v_{LP-} - v_{LP+}) := -Gv_d$ and $i_p = G(v_p - v_{REF}) = Gv_p$. The voltage on the capacitor is $v_p = -(i_1 + i_p)/sC_p$. Inserting the expression for the currents, and solving for the transfer function, we get

$$\frac{v_p}{v_d} = \frac{1}{1 + sC_p/G} \tag{1}$$

Two observations can be made at this point. In the first place, the DC value of this function is 1 (this result presumes perfect matching of the transconductances), so a true copy of the peak voltage is effectively stored on C_p. In the second place, this is a first-order LP function, so the voltage on the capacitor approaches the final value with no overshoot. This last fact is the justification for omitting the diode-connected transistor M_1 in the analysis. The result for the negative peak voltage is similar. The time constants C_p/G and C_n/G should be set in relation to the length of the start symbol, to allow the positive and negative peak voltages time to settle.

The digital *data* signal is extracted from the input signal by comparing it to the average of the two peak values. This is done by summing the currents $I_3 = -2GV_d$, $I_p = G(V_p - V_{REF})$ and $I_n = G(V_n - V_{REF})$ at the output node, and letting it slew to the positive supply or to ground according to the sign of the result.

The reset signal is used to initialize the values of the capacitors which store the positive and negative peak signal voltages. The positive peak voltage V_p is set to V_{SS} by the reset signal, while the negative peak voltage V_n is set to approximately four gate-source voltages. This last value is determined partly by a voltage clamp consisting of three diode-connected transistors. The purpose of the clamp is to limit the voltage swing on the summing node for the negative peak computation. This reduces the amplitude of the capacitive feed-through to C_n through the diode. No clamp is necessary for the positive peak since the available voltage range at the summing node is more limited. The negative peak reset transistor is not connected to V_{DD}, but to a current source, in order to limit the current through M_2 and the clamp during reset.

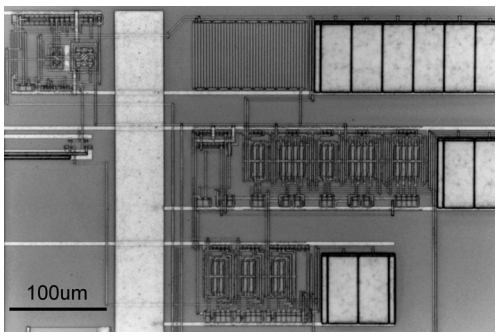


Figure 6. A photo of the demodulator core.

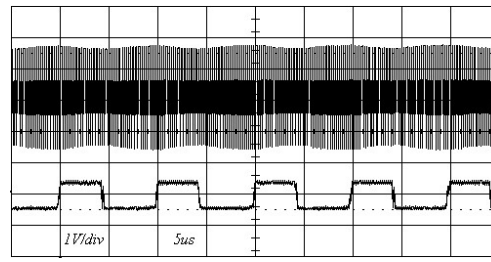


Figure 7. Measurement results for a 10 MHz carrier modulated 10% with a 0101... bit stream at 200kbit/s. The upper trace is the RF signal and the lower trace is the extracted data.

3. Experimental results

The demodulator was implemented in a standard digital 0.5 µm CMOS process, with one poly and three metal layers. The maximum supply voltage is 3.3 V.

All measurement results were obtained with a supply voltage of 3 V. In order to test the circuit with a wide range of carrier frequencies, modulation parameters and signal levels, a signal generator was used for most measurements. The functionality of the circuit was however verified with transmission through an inductive link with a carrier frequency of 5 MHz at 200kbit/s.

Due to the nature of the dual power/data link, the peak-to-peak amplitude of the unmodulated signal is almost constant, and equal to the supply voltage plus two diode voltage drops. In the following test results, the unmodulated signal level is fixed at $3.5\ \mathrm{V_{p\text{-}p}}$.

3.1. Carrier extraction

The sensitivity of the carrier extraction mechanism is limited by the parasitic capacitance on the *clk* signal output node in the squarer (figure 2). The necessary signal level for detection rises as the frequency increases, since the current into the output node is proportional to the signal. An output buffer was connected to the *clk* signal, to drive the signal off-chip for testing purposes. Due to a layout error, the parasitic capacitance on this node is higher than necessary, and the sensitivity of the carrier extraction does therefore fall off faster with frequency than expected.

The measurements show that the detection threshold is approximately $0.4\ \mathrm{V_{p\text{-}p}}$ up to 5 MHz, and rises to $3.6\ \mathrm{V_{p\text{-}p}}$ at 12 MHz. Simulations of the circuit without the added parasitic capacitance show a threshold of $0.8\ \mathrm{V_{p\text{-}p}}$ at 15 MHz.

3.2. Detection

The circuit works as expected for carrier frequencies between 1 and 15 MHz, with 10% AM modulation rates up to 200kbit/s. Figure 7 show the results for a 10 MHz carrier modulated at the maximum bit rate.

3.3. Power consumption

The supply current of the demodulator was measured for a wide range of operating conditions. At the nominal input signal level of $3.5\ \mathrm{V_{p\text{-}p}}$ the supply current was constant at 20.0 ± 0.5 µA over the carrier frequency range of 1–15 MHz. A weak dependence on the signal amplitude was found, with the supply current decreasing by a few percent as the amplitude was reduced to zero. This dependence is due to the class AB operation of the squarer circuit.

4. Conclusion

A demodulator circuit has been presented, which in addition to a low power consumption of 60 µW, is versatile enough to be used in a wide range of implantable devices. Measurements show that the demodulator can handle up to 200kbit/s with a carrier frequency of up to 15 MHz. The sensitivity of the circuit makes it suitable for systems which must operate with low modulation indexes, either due to power transfer constraints or transmitter design.

The demodulator is designed for intermittent data transmissions and does not rely on constant modulation to determine the modulation levels. This allows maximum power transmission between data sequences.

The data rate and carrier frequency range can be extended by modifying the cutoff frequency of the LPF, and the passive component values in the input network.

[1] D. Galbraith, M. Soma and R. White, "Wide-band efficient inductive transdermal power and data link with coupling insensitive gain", *IEEE Trans. BME*, **34** (1987), pp. 265–275.

[2] A. Djemouai, M. Sawan and M. Slamani, "An efficient RF power transfer and bidirectional data transmission to implantable electronic devices", *Proc. ISCAS '99*, pp. II-259.

[3] G. Nebel *et al.*, "ASK 10% demodulator for contactless smart card IC", *Proc. ESSCIRC '99*, pp. 230–232.

[4] O. Oliaei and P. Loumeau "Four-quadrant class AB CMOS current multiplier", *Electronics Letters*, **32** (1996), pp. 2327–2329.

A distributed transducer system for functional electrical stimulation

Gunnar Gudnason, Jannik H. Nielsen and Erik Bruun
Ørsted-DTU, Technical University of Denmark, DK-2800 Lyngby
gg@oersted.dtu.dk

Morten Haugland
Center for Sensory-Motor Interaction, Aalborg University, DK-9220 Aalborg

Abstract

Implanted transducers for functional electrical stimulation (FES) powered by inductive links are subject to conflicting requirements arising from low link efficiency, a low power budget and the need for protection of the weak signals against strong RF electromagnetic fields.

We propose a solution to these problems by partitioning the RF transceiver and sensor/actuator functions onto separate integrated circuits. By amplifying measured neural signals directly at the measurements site and converting them into the digital domain before passing them to the transceiver, the signal integrity is less likely to be affected by the inductive link. Neural stimulators are affected to a lesser degree, but still benefit from the partitioning.

As a test case, we have designed a transceiver and a sensor chip which implement this partitioning policy. The transceiver is designed to operate in the 6.78 MHz ISM band and consumes approximately 360µW. Both chips were implemented in a standard 0.5 µm CMOS technology, and use a 3 V supply voltage.

1. Introduction

The subject of this article is a solution for several problems which plague designers of inductively powered biomedical implants for functional electrical stimulation. Inductive links are commonly employed to power and control such implants since they provide a wireless connection, which is desirable since percutaneous wires provide an infection path into the body. Inductive links also eliminate the need for an implanted battery, which might eventually need to be replaced. The price which must be paid for these advantages is the relatively high power dissipation in the external (non-implanted) apparatus due to the low efficiency of the link, and a degradation of the signals under examination, because of the high electromagnetic fields. While the RF signals normally lie far outside the biological signal band, they can easily desensitize an amplifier designed for a 10–20µV signal range.

The signal amplitude quoted above is typical for signals obtained using cuff electrodes [1]. Stimulators connected to cuff electrodes generate much larger signals, so they are not affected to the same extent as sensors by the external RF field.

We propose therefore a physical partitioning of the signal processing functions, in order to limit the transmission distances for sensitive signals, and to allow optimal placement of critical functions of the system. The conflicting design criteria which can be accommodated to a large extent by partitioning are:

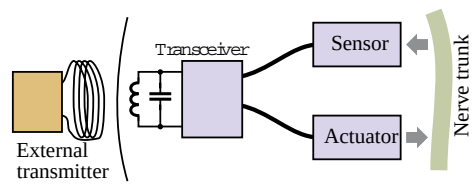


Figure 1. A partitioning example for a FES system containing a transceiver chip and sensor and actuator chips. The last two are not necessarily connected to the same nerve trunk.

- The RF part of the system should be close to the skin surface for better power transmission, and also for better link bandwidth. The proximity may also be an advantage for surgical access.
- The transducers should be placed close to the active sites, which can be relatively deep inside the body.
- Long routing of transducer signals should be avoided, as the strong electromagnetic field will induce an RF interferer overlaid on the desired signal. Physical separation of the transducer from the RF link also reduces the problems, since the field falls off rapidly with distance (as $1/r^3$ in the far-field limit).
- The system must contain a considerable amount of digital logic, for control, timing, and buffering of data. The logic will inevitably couple switching noise onto the supplies and into the substrate. Using low-noise logic solutions like current-steering logic [3] can eliminate the problem, but the static supply current of CSL and other ultra-low noise logic families makes them unsuitable as the system complexity passes a certain point. Placing the sensors on separate chips isolates them from the logic supply noise.

One option which was examined, and which alleviates some of the problems above, is to place only the receiver

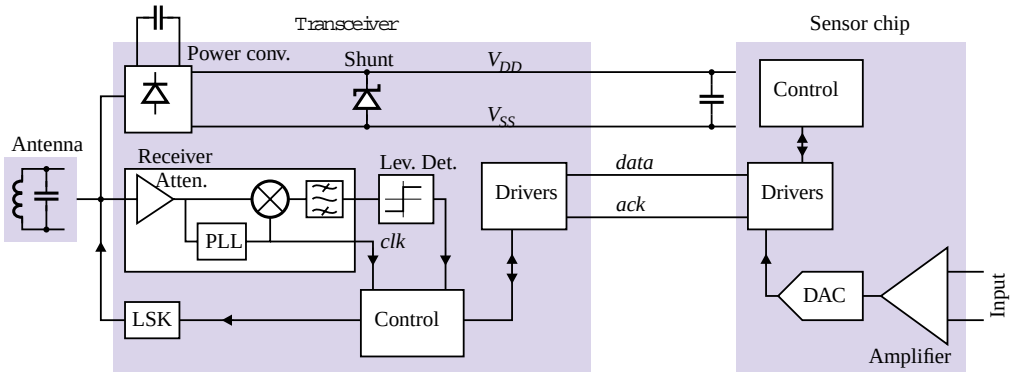


Figure 2. A simplified block diagram of the system, showing the main parts of the transceiver and sensor chips. Differential signals are shown with a single line for clarity.

coil close to the skin surface, and route the antenna signal to a single integrated transceiver/transducer. This is however not practical since biocompatible interconnect solutions like Cooper cable[1] have a high impedance at DC and RF.

We adopted a solution with simple digital communication protocols between the central transceiver and control chip, and the outlying transducer chips. The protocol was designed so that it could be accommodated along with the power supply wires in a simple 4-wire flexible cable. In this specific test case, we implemented a simple system with one transceiver chip and one sensor chip. The sensor chip includes an amplifier for cuff electrode signals, which is itself the subject of another article [2], an AD converter and bus interface logic. The transceiver includes a direct conversion receiver (actually a homodyne), a load modulation circuit for transmission of data out of the system, control logic, supply regulators and references.

2. System description

The following is a description of the transceiver chip and the aspects relating to communication between it and the other parts of the system. Since the main focus of this article is on the way in which partitioning of functions can solve some of the problems in an implanted system of this type, the sensor chip will only be described inasmuch as it relates to the top-level design of the system. The principle of partitioning applies to any transducer type that is relevant in an implanted system, so the internals of the sensor chip will not be emphasized. Figure 2 shows a top-level diagram of the system.

2.1. The link

The use of inductive links for power and data transmission is described in great detail elsewhere, see for example [5, 6].

By separating the transceiver chip and the antenna, it is possible to reduce the distance between the external transmitter antenna (a tuned LC circuit) and the internal one to 15–20mm. The coupling coefficient for normal coil geometries at this distance is on the order of 0.01–0.05. This permits far better power transfer than if the transceiver were placed together with the transducer, and makes load modulation a viable method to extract information from the system. Systems with smaller coupling coefficients must resort to active transmission of signals to the outside [4], which increases the power consumption.

The chosen modulation method is PAM with Manchester encoding and a modulation index of approximately 0.2. This is compatible with high-efficiency class D or E transmitter configurations [7, 8]. The carrier frequency is 6.78 MHz which coincides with one of the ISM (industrial, scientific and medical) bands. The target bit rate is 50kbit/s.

2.2. The power supply

The power is extracted from the RF carrier by means of a full-wave bridge rectifier, which is implemented by using diode-connected P-channel MOS transistors in a common N-well, as shown in figure 3. The standard CMOS process does not offer high-quality floating diodes, so another solution must be chosen. The available elements are three types of p-n junction and diode-connected MOS transistors, but these all suffer from high substrate currents and/or parasitic elements which divert some of the current from its intended path. The PMOS rectifier bridge is accompanied by parasitic vertical PNP transistors which can divert some of the input current to V_{SS} instead of to V_{DD}.

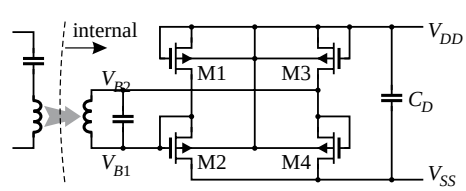


Figure 3. The power conversion circuit.

[1] Produced by Finetech Medical Ltd.

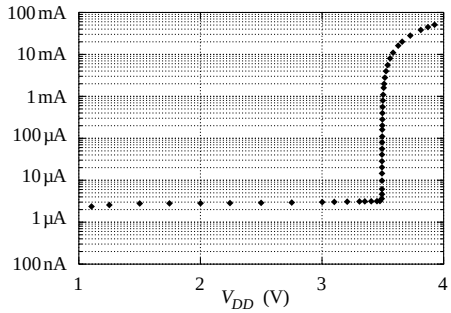


Figure 4. The measured shunt regulator current.

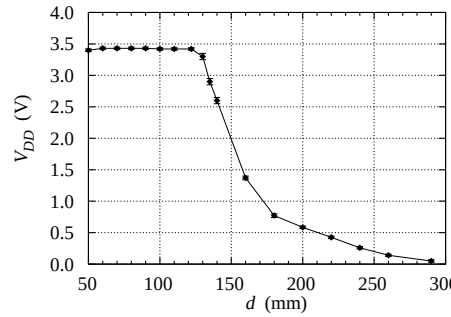


Figure 5. The measured on-chip supply voltage as a function of the separation between the transmitter and transceiver antennas.

We have solved this problem by dimensioning the PMOS transistors so that they are biased in weak inversion over the entire operating range. Since the nominal threshold voltage is about 0.6 V, there is a range of drain-gate voltages for M1 in figure 3 (corresponding to the emitter-base potential in the parasitic PNP) where the MOS current dominates the current through the bipolar transistor by a large factor. The current-handling capability of the rectifiers can be increased for a given maximum parasitic/main current ratio, by increasing the width (and area) of the transistors. The main penalty is an increased input capacitance. Our experimental data show that the parasitic current ratio is smaller than 0.005 for a system supply current of 200 µA and an input capacitance of 2 pF. All the transistors in figure 3 have $W/L = 800/0.5$ in micrometers.

The transducer chip contains an active shunt regulator which provides protection against excessive input power. The transmitted power is a very strong function of antenna spacing, so an unprotected circuit can easily be burned out. The shunt circuit is basically a pass transistor between V_{DD} and V_{SS}, and a feedback loop which compares V_{DD} to a bandgap reference voltage. As the supply voltage approaches the trigger point (which was set equal to the maximum supply voltage for the technology), the regulator begins shunting current from the supply, and the current changes by 4 decades over a short supply voltage interval (see figure 4).

The performance of the power supply conditioning system is shown in figure 5. The conversion of the RF power to a DC supply voltage and the overvoltage protection is carried out mostly on-chip, with the only external component being the energy storage capacitor C_D. The transmitter used in these measurements was running at a relatively high power level, witnessed by the fact that the full supply voltage is reached at a separation of 120 mm. Despite the strong dependence of the magnetic field amplitude on the separation d, the on-chip shunt succeeds in maintaining the supply voltage at the nominal level at all separations.

The transceiver chip provides the supply voltage for the other chips in the system. Two strands out of four in the Cooper cable have been assigned to V_{SS} and V_{DD}. Because of the helically wound construction of the cable, and the Pt-Ir composition, the impedance is relatively high at all frequencies. To provide a stable supply voltage for the outlying chips, it is therefore necessary to add decoupling capacitors to the supply lines at the transducer ends. A stimulator chip can require relatively large current pulses from the supply while delivering a stimulus.

2.3. Transceiver

The receiver part of the transceiver chip is a direct-conversion receiver. Since the input signal has a larger amplitude than the supply range, an attenuator with a gain of approximately 0.1 is inserted in the signal path. The input signal is tracked by a PLL which also provides the system clock (the block diagram does not show a clock divider circuit which divides the clock frequency down to 1.7 MHz).

The output of the PLL and the input signal are put through a mixer, and the mixer output is filtered in a 4th-order differential G_m–C filter. The filter output is directed to a mixed analog-digital level detector, which detects the sign of the transmitted data bits. Instead of using a partly analog level detector, the normal processing method would be to sample the filter output for further digital signal processing. This does however require more sophisticated digital circuits than we were willing to implement, and possibly requires more supply current.

The target bit rate for the system was 50kbit/s, and the receiver is designed to handle up to 100kbit/s.

Data is transmitted out of the system by load modulation. The reflected impedance seen by the external transmitter is varied by changing the load seen by the secondary LC circuit. A switch is connected between the terminals of the antenna, and by closing the switch, a maximum change in the reflected impedance is obtained. This simple scheme has the disadvantage of stopping power transfer to the system during load modulation, and other load modulation circuits have been designed to avoid this [9]. The duty cycle of the switch closure is however so low in our case that the reduction in average power transfer is small.

2.4. Interchip communication

The cable type that we used as a reference for our interconnects is, as mentioned before, a 4-strand biocompati-

ble cable type by the name of Cooper cable. In addition to being biocompatible, the cable is wound in a helical pattern and embedded in silicone, so it can be stretched. The stretching reduces the probability of damage to the surrounding tissue.

Because of the biological constraints that the cable must fulfill, it does not have very good electrical properties, and the use of the cable must be adjusted accordingly. We measured the series resistance at DC of a representative length of the cable, and found it to be 210 Ω/m. The capacitance between any two wires varies from 20–80 pF/m because of the lack of symmetry, and the capacitance of each wire to the surroundings is about 50 pF/m.

No clock signal is sent across the connection between the chips. Instead, we use an asynchronous handshake mechanism to control data transfer, and an internal oscillator in the sensor chip to provide a time base.

The line drivers are class AB circuits which have a quiescent current consumption of 5 µA each, and can slew the line voltages with a 50 µA current. This type of driver was chosen instead of faster types because this is more than sufficient for the purpose, and by limiting the slewing currents the supply transients are reduced. The high and low voltages on the signal lines are 1.0 V and 0.3 V respectively instead of the full supply range, again to reduce supply transients and power consumption.

3. Measurement results

Most of the basic functions of the transceiver and sensor chips have been measured, and they behave according to specifications.

The results for the power conversion circuit show that the use of PMOS transistors in weak inversion as rectifiers is an ideal solution for pure CMOS technologies. Current throught the substrate are eliminated by placing everything in an N-well, and the effect of parasitic bipolars are all but eliminated.

The active shunt regulator and other regulators and references on the chip are up to the design criteria. Specifically, the shunt regulator consumes negligible supply current within the normal voltage range, with a very sharp rise in current as the trigger point is exceeded. Previous systems use passive zener-based regulators, whose gradual I-V characteristic provides insufficient overvoltage protection for low-voltage CMOS technologies.

The data link to the transceiver chip was tested by using a Class D transmitter driving the inductive link, with a data rate of 50kbit/s and 20% ASK modulation and Manchester encoding. The data transfer functions according to the specifications, and higher data rates can easily be supported with minor modifications.

4. Conclusion

We have demonstrated a partitioning scheme for implanted sensor and actuator devices which places the signal processing functions where they are needed. By using this scheme, it is possible to provide better isolation of weak biological signals from strong external disturbances, while simultaneously reducing the overall power consumption by placing the transceiver closer to the external interface. The communication between separate parts of the system has been adapted to existing biocompatible interconnect methods.

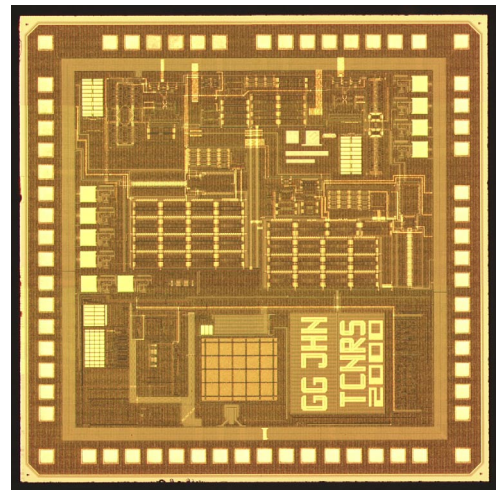


Figure 6. Chip die microphotograph.

We have implemented a simple system consisting of a transceiver and a single sensor chip, but the concept can easily be extended to more general combinations of sensors and actuators, in order to created a complete neural stimulation system with a closed feedback loop.

[1] G. E. Loeb and R. A. Peck, "Cuff electrodes for chronic stimulation and recording of peripheral nerve activity". *J. Neuroscience Methods*, **64** (1996), 95–103.

[2] J. H. Nielsen and T. Lehmann, "An implantable CMOS amplifier for nerve signals", *Proc. ICECS 2001*.

[3] H.-T. Ng and D. J. Allstot, "CMOS current-steering logic for low-voltage mixed-signal integrated circuits", *IEEE Trans. VLSI*, **5** (1997), pp. 301–308.

[4] M. Nardin and K. Najafi "A multichannel neuromuscular microstimulator with bi-directional telemetry", *1995 Conf. on Solid-State Sensors and Actuators*, pp. 59–62.

[5] D. Galbraith, M. Soma and R. White, "Wide-band efficient inductive transdermal power and data link with coupling insensitive gain", *IEEE Trans. BME*, **34** (1987), pp. 265–275.

[6] N. de N. Donaldson and T. A. Perkins. "Analysis of resonant coupled coils in the design of radio frequency transcutaneous links". *Med. and Bio. Eng. and Computing*, **21** (1983), pp. 612–627.

[7] A. Djemouai, M. Sawan and M. Slamani, "An efficient RF power transfer and bidirectional data transmission to implantable electronic devices", *Proc. ISCAS '99*, pp. II-259.

[8] P. R. Troyk and M. A. K. Schwan, "Closed-loop class E transcutaneous power and data link for microimplants" , *Trans. BME*, **39** (1992), pp. 589-599.

[9] Z. Tang et al. "Data transmission from an implantable biotelemeter by load-shift keying using a circuit configuration modulator". *IEEE Trans. BME*, **42** (1995), pp. 524.